The Dissertation Committee for Grant Michael Gelderman Certifies that this is the approved version of the following dissertation:

Developing novel tools to characterize regulatory RNA-protein networks *in vivo*

Committee:

Lydia Contreras, Supervisor

Hal Alper

George Georgiou

Andrew Ellington

Shelley Payne

Developing novel tools to characterize regulatory RNA-protein networks *in vivo*

by

Grant Michael Gelderman, B.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2016

Developing novel tools to characterize regulatory RNA-protein networks *in vivo*

Grant Michael Gelderman, Ph.D.

The University of Texas at Austin, 2016

Supervisor: Lydia Contreras

This work focuses on the design and implementation of tools to characterize a regulatory RNA-protein network *in vivo*. As such, this work developed techniques to detect RNA-protein interactions, characterize repression of translation, and model the RNA-protein interaction as a means of target identification. We specifically focused on the RNA-protein interactions of the post transcriptional regulatory protein, CsrA, of the Carbon storage regulator system.

A fluorescence based technique, named Tri-molecular Fluorescent Complementation (TriFC), was developed to detect RNA-protein interactions *in vivo* and demonstrated using CsrA and the sRNA, CsrB. This tool showed sensitivity to mutations in CsrA that affect the affinity of the interaction. A modified version of this tool was later implemented to characterize mRNA-CsrA interactions as a means of identifying targets of CsrA.

A translational assay was developed to detect the effect that CsrA has on the translational activity of mRNAs. This assay tested 241 different mRNA targets and identified 94 mRNAs that displayed inhibition by CsrA. Using modifications to the TriFC system, 32 mRNAs were observed directly interacting with CsrA. The evidence from the TriFC system and the translational assay revealed 19 previously uncharacterized

mRNAs as targets of CsrA regulation, with the most important of these genes being the sigma factor *rpoS*.

Lastly, a model was developed using the primary principles of CsrA interactions that predicts the affinity an mRNA sequence has for CsrA and the effect that CsrA would have on the translational activity of the mRNA. Experimental data established that mRNA targets with a sufficiently a high prediction of CsrA affinity are very likely to be true CsrA targets. The predictions for the binding site affinity were evaluated using a novel *in vivo* titration technique, and it was demonstrated that the affinity predictions have physical significance in describing the interaction of CsrA and the RNA. With this model, it will be possible to evaluate the genome of *E. coli* to predict CsrA's regulatory effect on mRNA targets. In total, this work demonstrated generalizable tools to characterize RNA-protein interactions, and the specific focus on the CsrA protein makes these tools immediately useful for the characterization of this significant regulatory system.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction and Background Information

1.1 Introduction

The recognized role of cellular RNAs has greatly expanded over the years to include a wide diversity of regulatory actions. In bacteria, small RNAs (sRNAs) regulate outer membrane proteins (1), quorum sensing (2, 3), expression of virulence-related genes (4), response to stress (5), and general control of metabolism (6). While there have been over 1,000 sRNAs identified (7), only a few mechanisms are known in detail. For instance, it is understood that sRNAs can function by 1) base-pairing with mRNAs to prevent protein translation and promote degradation, 2) sequestering proteins into ribonucleoprotein complexes to prevent their activity, or 3) directly catalyzing mRNA and protein degradation (see for review: (8, 9)). Additionally, RNA-based regulation employs modular recognition sequences that have the capability of binding to many different targets, creating a degree global regulatory control.

The role of global regulation is important for the response and survival of cells in new and changing environments. These transitions require the rewiring of expression and simultaneous coordination of potentially hundreds of genes for a rapid and robust response to stimuli (e.g. nutrient availability, osmolarity, pH, and temperature (5, 8, 10-14)).

In order to be an effective global regulator, an RNA or a protein needs to be able to recognize and differentiate the genes in their network from other genes. This

need is met through the use of a genetic barcode found that is recognized by their regulator. For RNA binding proteins, this barcode takes the form of a sequence motif that physically interacts with the protein (15).

An identified motif can accelerate regulatory network characterization by informing the creation of a model to predict targets that interact with the protein. However, significant experimental effort is still required to find the true extent of the regulatory network. Additional effort is frequently required due to the fact that the context of the motif (i.e. secondary structure, position of motif in respect to promoter or untranslated region, etc.) holds importance to the degree of regulation and the effectiveness of the model to make accurate predictions.

1.2 THE CARBON STORAGE REGULATOR: A GLOBAL RNP REGULATOR

One regulatory system of particular interest that relies on RNA-protein interactions is the Carbon storage regulator (Csr) system from *Escherichia coli*. This widely conserved global regulatory system exhibits control over hundreds of genes through an interaction based on a simple RNA motif (16). In *E. coli*, there are four important components of the Csr system: two proteins, CsrA and CsrD, and two sRNAs, CsrB and CsrC (17-19). The central component of the system is the CsrA protein. CsrA has an important role in cellular regulation, as it has been proven to be conditionally essential in cells grown in glycolytic media (20). The CsrA protein is composed of only 61 amino acids and forms a homodimer capable of binding RNA, specifically mRNAs. The interaction between the CsrA protein and a target mRNA affects the translation activity or stability of the mRNA (21). CsrB and CsrC interfere with the CsrA-mRNA

interaction by binding CsrA and competing with the mRNAs for the available CsrA. The abundance of CsrB and CsrC is modulated by the protein CsrD which directs degradation of the sRNAs (19, 22-28).

CsrA recognizes its RNA targets through the nucleotide motif GGA (29). The GGA motif was first identified by its repetitive presence in CsrB (30). Specifically, SELEX studies have shown that CsrA preferentially binds to AAGGA motifs or ANGGA motifs located in the loop of a hairpin (29, 31, 32). While the hexaloop structure is the preferred binding site of CsrA, NMR studies of RsmE, a closely related homologue of CsrA in *Pseudomonas aeruginosa*, have physically described the interactions between RsmE and lower preference variants of the GGA motifs and demonstrated that the hairpin structure is not necessary for the interaction (33). Furthermore, CsrA is capable of binding to two RNA sites and favors binding sites with a minimum distance of 10 nucleotides between them (32).

Interestingly, CsrA does not appear to have any catalytic activity, and the only purpose of CsrA is to bind to RNA. Characterizations of mRNA-CsrA interactions have revealed the mechanism that CsrA uses to regulate mRNA activity. Early observations of CsrA binding to mRNAs found that CsrA predominantly bound to GGA motifs in the 5' UTRs of transcripts (19, 34). More extensive RNA footprinting of CsrA targets, such as *glgC, pgaA, sdiA, cstA,* and *hfq* (19, 34-37), have shown that CsrA binds multiple locations near the ribosome binding site of the mRNA. Considering that the CsrA recognition sequence, GGA, is similar to the Shine-Delgarno sequence, it is clear that CsrA functions by binding the mRNA to prevent the ribosome from binding.

While the most common mechanism of CsrA regulation is by interfering with ribosome binding, CsrA also utilizes other unique mechanisms to control expression of its targets. CsrA binding has been shown to inhibit gene expression by altering the structure of the mRNA so it becomes susceptible to RNAse degradation (34). Conversely, CsrA has been known to activate translation by altering the mRNA structure to favor ribosome binding (38) or to block rho utilization sites and stabilize the mRNA (39). CsrA interactions have also been shown to rely on post-transcriptional modifications (40). Interestingly, CsrA binding site analysis has suggested that CsrA can bind to any region of an mRNA (41), but all of the currently understood mechanisms of CsrA regulation require CsrA binding in the 5'UTR or near the translation initiation.

With such a simple mechanism at the core of CsrA inhibition, the important question concerning CsrA is how it is capable of preferentially binding and regulating specific mRNA targets when the Shine-Delgarno is so similar to the CsrA site. One suggestion is that CsrA's two RNA binding sites act cooperatively; one high affinity RNA site attracts CsrA so that the other binding site will be attracted to the RBS (32). However, this model is likely incomplete; the CsrA targets *hfq* and *ycdT* contain only a single GGA motif (36, 42). The mechanism that CsrA employs to preferentially bind these targets is not yet clear and raises concerns over the extent that CsrA regulates all mRNAs.

Several experimental studies have been performed to measure the impact of CsrA on cellular pathways in *E. coli*. In these studies, genes impacted by the Csr system were identified using CsrA pulldowns (41, 43), microarrays (42), proteomics (21), C13 flux

analysis (44), and transcriptomic analysis of mRNA degradation patterns (45). Combined, these studies suggest that CsrA has an effect on 200-800 genes. That estimate indicates a large degree of uncertainty and calls into question whether CsrA interacts with these genes directly through the mRNAs or has a broader and indirect effect on these genes.

Several efforts have been made to predict targets of CsrA by searching the genome for GGA motif variants. The first attempt to predict CsrA targets searched the genome for 3 degenerate versions of motifs containing GGA found within a 25 nucleotide window of the translation initiation site for all *E. coli* genes (21). This method predicted about 700 mRNAs as potential targets of CsrA. An alternative prediction model identified targets based on the number of GGA motifs present in 5'UTR (46). A positive prediction of a CsrA target required that multiple GGA sites be within 60 nucleotides of each other and present within the 5' UTR of a gene CsrA target. These constraints identified approximately 160 potential targets in *E. coli* (46). These two sets of predictions vary widely in both the number and identity of the targets that they predicted. Different interpretations of the CsrA binding interaction will significantly impact computational predictions of mRNA targets.

With the ambiguity in omics based measurements and predictive models, there has been slow progress in identifying the true, direct targets of CsrA. Considering the high number of potential CsrA-mRNA relationships, only about 17 targets have been well characterized experimentally in the literature. To rectify the dearth of knowledge surrounding such an important regulator, this work focused on developing methods

and models to better understand the nature of the RNA-protein interactions of the Csr system.

1.3 Experimental Characterization of RNA-Protein Interactions

The identification of an sRNA's protein target has been a difficult task, as typical detection of interacting sRNA-protein pairs often requires extensive experimental characterization (47). Indeed, only two examples of regulatory sRNA-protein interactions have been identified and extensively characterized in bacteria: the Csr system (24, 25) and the 6S RNA (48). However, recent advancements in RNA sequencing technologies have made largescale discovery type of experiments more approachable (49).

One of the major techniques for the largescale characterization of RNP interactions is cross-linking and immunoprecipitation (CLIP) (50). In CLIP procedures, cells are subjected to cross linking conditions (such as formaldehyde or UV light) to covalently link RNAs to proteins. The RNA-protein complex (RNP) of interest is purified from cell lysate using immunoprecipitation techniques, and the bound RNAs are identified using RNA-seq technologies. While variations on CLIP technologies have sought to improve the specificity of target interactions and identify exact binding sequences (41, 51), CLIP technologies are often criticized for being too promiscuous and identifying superfluous RNA-protein interactions or for requiring too high of an affinity of the RNA-protein interaction for accurate detection (52).

In addition to the limitations in identifying sRNA-protein interactions, current methods for studying and characterizing RNA-protein interactions are largely limited

to *in vitro* studies such as gel mobility shift assays (53), SELEX (54), electron microscopy (55), X-ray diffraction, and SAXS (56). While these tools are highly valuable, they fail to capture the complex natural roles RNPs play within their native cellular context.

Current technologies available to study RNP interactions *in vivo* largely focus around fluorescence based techniques. Three common techniques are FRET, Yeast Three Hybrid systems, and fluorescence complementation techniques. FRET stands for Fluorescence Resonance Energy Transfer and is a common technique used to demonstrate when two molecules are in close proximity to each other. The concept behind FRET is that a fluorescence signal is produced when two fluorophores are in close proximity and the excitation of one fluorophore excites the second fluorophore to emit a unique fluorescence signal (57). The need for fluorophores can make *in vivo* characterization difficult because additional chemical modification is required to covalently link a fluorophore to the molecule of interest. FRET has been applied using *in vivo* produced fluorescent proteins, but these systems have been limited to eukaryotic systems and evaluation of protein-protein interactions (58).

The Three Hybrid Systems creates a fluorescence signal by using an RNA-protein interaction to initiate transcription of a fluorescent protein (59). This system employs three fusion molecules to detect the interaction between the protein and RNA of interest. The protein of interest is fused to a transcription factor, and the RNA of interest is fused to a protein binding motif. A third party protein fusion consisting of a DNA binding protein and an RNA binding protein tethers the RNA to

the DNA through the interaction of the RNA binding protein and the protein binding motif on the RNA fusion. When the protein of interest binds to the RNA of interest, the transcription factor is directed to the DNA to transcribe the signal for a fluorescent protein. While this system does provide a relatively easy to use detection method, this system has only been applied in yeast and is not suited for prokaryotic systems. Additionally, this system relies upon tethering the RNA to the DNA, which can remove the functional context of the RNA-protein interaction of interest.

The final fluorescence technique discussed here is the complementation assay. In a standard complementation assay, a fluorescent protein is divided into two fragments expressed as chimeric fusions to two separate proteins; this set up produces a fluorescence signal when the two proteins come into close proximity to allow refolding of the fluorescent protein fragments (60). Although the most common complementation assay examines protein-protein interactions in a technique described as Bi-molecular Fluorescence Complementation (BiFC), split protein complementation has been used to visualize mRNAs in bacterial and eukaryotic systems by tethering multiple fluorescent protein fragments to the RNA of interest by including the recognition sequences of well-known RNA binding proteins (61-63). This strategy of using third party RNA-protein interactions was referred to as Tri-molecular Fluorescence Complementation (TriFC) when it was first implemented to detect and characterize mRNA-protein interactions in mammalian cells (64). Even though this strategy can be negatively influenced by the interaction assuming an orientation that is unfavorable for the refolding of the fluorescent protein, this

technique is popular because it can be implemented easily using plasmid based expression systems. Since TriFC has been demonstrated in prokaryotes, does not require extensive chemical modification to the molecules of interest, and can be created using a plasmid system, TriFC is well suited for studying the RNA-protein interactions of the Csr system *in vivo*.

1.4 Model Characterization of RNA-Protein Interactions

Physical characterization may be the first step in understanding any regulatory interaction, but modern biology looks to improve the understanding of regulation by developing computational tools to model and predict cellular regulation. One of the great challenges in modern biology is to understand the extent that any particular regulatory interaction has on the phenotype of a cell. In order to achieve this task, research has focused on using computational methods to predict the genetic targets of regulatory proteins and RNAs. Currently, there is significant interest being paid to archiving omics data and model based predictions of regulatory interactions (65) with the hope that a more complete understanding of regulation can help develop better models to predict transcriptional outcomes (66). While identification of mRNA targets of sRNAs has been aided by genetic and bioinformatics analyses built around the understanding of nucleotide base pair interactions (67), there are no such generalizable models to predict DNA-protein or RNA-protein interactions. There are models being developed that attempt to predict RNA-protein behavior based on principles of atomic structure (68, 69), but these models can be difficult to implement because they require knowledge of the protein crystal structure. In general, predictive and mechanistic models of nucleotide-protein interactions are exceptionally limited and require experimental verification of the protein's binding motif or consensus sequence.

While a variety of free energy models have been developed to describe transcription factors binding to DNA (70-72), free energy interpretations of RNA-protein interactions are typically quite rare. This discrepancy in model availability is largely related to the historical focus on transcription initiation and the general underappreciation of post-transcriptional regulation. Fortunately, the principles used to make DNA-protein interaction models can be applied to RNA-protein interactions.

For DNA binding proteins, typical model development requires the identification of a consensus binding motif which is typically used to develop a position weight matrix or a free energy of binding model. Simple models rely on the sequence similarity to the recognized motif in order to make a prediction on whether a sequence is a target of a protein, but more advanced models can use the sequence similarity to predict the affinity that the sequence has with the protein. These more advanced models require a reliable position weight matrix or free energy model that accounts for a nucleotide's individual contribution to the binding interaction. The estimation a nucleotide's individual contribution requires either experimental data, such as SELEX and other mutational studies, or predictions using Molecular Dynamics simulations (73). While developing an accurate model for nucleotide-protein interactions requires extensive characterization, a reliable model greatly aids the development of regulatory network models by identifying uncharacterized and unrecognized interactions that would otherwise be missed through experimental techniques.

1.5 Summary of Research Objectives and Accomplishments

Characterization of the CsrA-RNA interaction through experimental and model based techniques is an interesting technical challenge due to the overall number and complexity of the interactions. For example, the protein CsrA forms a dimer to bind

RNA at two binding interfaces; CsrB, for its part, is known to bind approximately nine CsrA dimers (30). This redundancy in binding sites suggests that molecular geometry plays a critical role in CsrA-RNA recognition. There is also a redundant sRNA (CsrC) that appears to duplicate CsrB function. Finally, CsrA is suspected to interact with hundreds of mRNAs. All of this suggests that CsrA is highly tuned to elicit various modes of regulation specific to the conditions of the cell.

In order to better understand CsrA-RNA interactions, this work develops two fluorescence based assays to study CsrA interactions. One assay detects the interaction of the RNA with CsrA, and the other assay provides evidence of the regulatory effect of the interactions. Additionally, this work creates a model to predict the likelihood of CsrA regulating a specific mRNA.

Chapter 2 of this dissertation describes efforts to create an *in vivo* reporter of RNA-protein interactions using a fluorescence complementation assay. Specifically, we developed the Tri-molecular Fluorescence Complementation assay (TriFC, Section 1.3) to detect the interaction of the protein, CsrA, and its regulatory sRNA, CsrB. In this assay, the protein CsrA was fused to the N-terminal segment of YFP. To direct the C-terminal domain to CsrB, CsrB was modified to include the binding sites of the well-documented RNA binding protein of the MS2 bacteriophage coat protein. The C-terminal complement of YFP is fused to the MS2 coat protein. When the two protein fusions bind to the modified CsrB RNA, we observed a fluorescence signal. The chapter discusses the practical design efforts that went into creating the fusion molecules. Importantly, we showed the ability of this fluorescence-based system to capture a range

of interactions from CsrA mutants that display differential affinity to an RNA target (74). This quantitative nature to a three partite complementation assay had previously not been demonstrated in the context of detecting regulatory sRNA-protein interactions. This work also served as the basis for an assay to detect specific mRNA-CsrA interactions (discussed in Chapter 3).

In Chapter 3, we investigated the interaction that CsrA has with 241 different mRNA targets using a translation assay. The 241 potential targets were selected using information from transcriptomics, proteomics, and CsrA co-purification assays. This translational fusion assay yielded a fluorescence signal that was related to the stability and translational activity of the mRNA targeted. In this system, CsrA concentration was modified through an inducible promoter, and variations in the fluorescence output were used as evidence of CsrA acting on the tested mRNA. Target mRNAs that displayed strong translation dependence to CsrA were evaluated using TriFC to detect the direct interaction between CsrA and the mRNA. With the two assays, we identified 19 previously unrecognized targets of CsrA regulation.

Chapter 4 discusses the creation of a CsrA-RNA binding model that predicts the effects of CsrA binding on translation. The effort to model CsrA targets used sequence information to generate a free energy based model to predict the affinity CsrA has to specific sites on mRNA targets. The model evaluates the impact of potential binding sites by generating separate predictions for CsrA binding to a single site and for CsrA binding multiple sites in the mRNA. Beyond predicting CsrA binding, the model predicts the regulatory effect of the interaction. To predict CsrA regulation, the model uses the

predicted secondary structure of the mRNA bound by CsrA as an input into the RBS Calculator which generates an overall prediction of the translation rate. The model of CsrA-mRNA interactions fits well into the RBS Calculator, as the prediction of CsrA binding is similar to how the RBS Calculator uses sequence based information to model mRNA-ribosome interactions (75). The predictions of CsrA regulation was compared to data generated from the 5'UTR translation assay (Chapter 3) to determine the effectiveness of the model in predicting mRNA-CsrA regulation. Future application of the model will predict CsrA interactions over the entire genome.

Chapter 2

Development of a Tri-molecular Fluorescent Complementation assay (TriFC) to detect regulatory RNA-protein interactions

This work has been previously published*. I was the primary author on this publication.

* Gelderman G, Sivakumar A, Lipp S, & Contreras L (2015) Adaptation of Tri-molecular fluorescence complementation allows assaying of regulatory Csr RNA-protein interactions in bacteria. *Biotechnology and Bioengineering* 112(2):365-375.

2.1 INTRODUCTION

One of the most basic needs of any biological characterization is the ability to detect interacting molecules. While *in vitro* techniques, such as electro-magnetic shift assays, are powerful methods for biophysical characterization, they are often criticized as lacking the proper cellular context for the interaction. This is especially true for sRNA based interactions due to the fact that most *in vitro* characterizations of sRNA rely upon *in vitro* transcription to produce sufficient quantities of sRNA for the assay. The preparation technique thereby removes any post-transcriptional processing that would normally be carried out in cellular conditions. These *in vitro* assays can be hindered by the misfolding or insolubility of the biological molecules which necessitates time consuming optimizations of the reaction conditions. Furthermore, these reaction conditions can be so extreme that they produce molecular orientations not relevant to *in vivo* conditions (RNA folding is particularly sensitive to magnesium ion concentrations). For these reasons, researchers often prefer *in vivo* techniques to detect interactions between biological molecules.

To detect most biological interactions, it is often necessary to engineer a method to produce a measurable signal in response to the desired interaction. An ideal system for the detection of inter-molecular interactions *in vivo* would be minimally invasive to the

cell, require minimal external equipment or reagents, and allow the interaction of interest to occur within normal biological conditions. Common techniques for the detection and characterization of interactions are Fluorescence Resonance Energy Transfer (FRET) (57), two and three hybrid systems (59), and complementation assays (76) (Section 1.3).

Since each of these techniques have their advantages and disadvantages, the context of the interaction under investigation is important. Our body of work is largely concerned with RNA-protein interactions for prokaryotic systems with specific focus on the interactions of CsrA, a global, post-transcriptional regulator known to bind many different RNA targets. FRET would not be an ideal system for RNA-protein interactions as it requires additional extracellular modification of the molecules of interest. Three Hybrid Systems shows potential to detect RNA-protein interactions, but this system has only been applied in yeast. Fortunately, molecular complementation has been demonstrated to function in prokaryotic systems and can be accomplished using low cost plasmid based cloning techniques.

In this study, the concepts of complementation assays are applied to detect the RNA-protein interaction of the protein CsrA and its sRNA, CsrB. Other studies have proposed TriFC in prokaryotic organisms before, using the technique to track the RNA's presence in the cells (61, 63, 77). Our work represents the first application of a complementation assay to study an RNA-protein interaction in a prokaryotic system.

2.2 RESULTS

2.2.1 An *in vivo* sRNA-protein Tri-molecular Fluorescence Complementation system to detect CsrB-CsrA interactions in *E. coli*

The system to interrogate the CsrB-CsrA (sRNA-protein) interaction was adapted from the Tri-molecular Fluorescence Complementation (TriFC) RNA visualization model (64). This system employs three molecular fusions (two protein

fusions and an RNA hybrid molecule) to allow visualization of RNA-protein interactions via the refolding of two inactive fragments of the enhanced yellow fluorescent protein (YFP). In this scheme, YFP is split into two inactive fragments, the N-terminal domain of YFP (NYFP; amino acids 1-154) and the C-terminal domain (CYFP; amino acids 155-238). These two fragments reassemble into an active, fluorescent protein when brought into close proximity by the interaction of the RNA and protein molecules. In our version of this scheme (illustrated in Figure 2.1A), CYFP is expressed as a fusion with the MS2 bacteriophage coat protein (MS2-CYFP), and NYFP is expressed as a fusion with the CsrA regulatory protein (CsrA-NYFP). The protein MS2 binds RNA with a high affinity for a specific RNA hairpin that is referred to here as the MS2 binding domain (MS2bd) (78). The two protein fusions are brought into close proximity by an RNA hybrid molecule containing the RNA sequences that bind CsrA and the MS2 coat protein, (CsrB and MS2bd, respectively). Each of these constructs is under promotional control from a distinct *lac* promoter. The assumption is that the interaction of CsrA fusion and CsrB fusion results in a measurable change in fluorescence that is mediated by the MS2bd-MS2 interaction.

To construct the MS2-CYFP fusion, the CYFP fragment was tethered to the dlFG variant of MS2 (79) using a previously developed linker (80). The construction of the CsrB-2MS2bd fusion was guided by applications of the MS2bd-MS2 interactions used for ribonucleoprotein purifications (81, 82). In these schemes, multiple repeats of MS2bd are used to attract the MS2 protein to RNA molecules. We expressed the MS2bd RNA as a series of two of the high affinity stem loops to create

the RNA segment called 2MS2bd. In addition, three of the previously mentioned segments were self-ligated to form a longer RNA tag containing six of the high affinity stem loops (6MS2bd). In this case, CsrB is expressed with either the 2MS2bd or 6MS2bd, both inserted at nucleotide 320 of CsrB. We reasoned that the native folding of the CsrB RNA would be less affected by the addition of a non-native sequence at the 3' end due to known patterns of co-transcriptional folding prevalent for intracellular RNA. The position at nucleotide 320 was chosen because available CsrB structural data predicted that the next structural element would be an intrinsic transcriptional terminator (30). Lastly, the CsrA-NYFP chimera was constructed using a flexible linker (3X(GGGGS)) to connect the CsrA and NYFP proteins. The linker sequence was intended to be long enough to overcome possible steric hindrance from the complex geometry of a CsrB-CsrA interaction that could prevent YFP refolding. The NYFP segment included a C-terminal FLAG epitope tag (DYKDDDDK) to allow for the assaying of protein expression. These three constructs were combined into two plasmids (Figure 2.1B) to reduce the metabolic burden of expressing three individual plasmids. The MS2-CYFP construct is expressed on the plasmid pMS2-CYFP, and the interacting RNA fusion and protein fusion of interest are both expressed from the plasmid pCsrA-NYFP+CsrB-MS2bd. For negative controls, plasmids were created that encoded only the CYFP, NYFP, or MS2bd sequence without the upstream fusion.

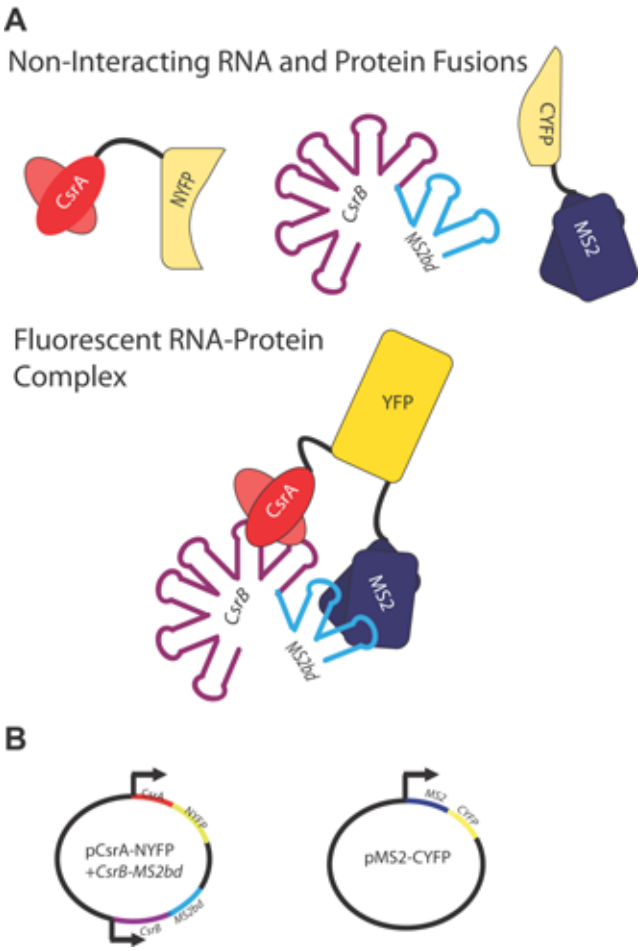


Figure 2.1: Tri-molecular fluorescence complementation method for studying Csr regulation

A. Illustrations of the non-fluorescent RNA and protein fusions (top) and fluorescent RNA-protein complex (bottom). The protein of interest, CsrA, is expressed as a protein fusion with the N-terminal domain of YFP (NYFP), yielding CsrA-NYFP. The C-terminal domain of YFP (CYFP) is expressed as a fusion to the MS2 bacteriophage coat protein (MS2), yielding MS2-CYFP. The hybrid RNA molecule contains the RNA of interest, CsrB, and the binding domains for the MS2 coat protein, MS2bd, and is called CsrB-MS2bd. The interaction of all three molecules results in fluorescence, as the interaction between the target protein and sRNA (CsrA and CsrB) drives the complementary fragments of YFP (NYFP and CYFP) to reassemble (bottom).

B. Depictions of the two plasmids that are used to encode the three fusion products necessary for fluorescent complementation. The protein and RNA of interest are expressed on the same plasmid, while the RNA tagging

molecule (MS2-CYFP) is expressed on its own plasmid. Each construct is under transcriptional control from a *lac* promoter.

2.2.2 The proteins and sRNA retain biological function when expressed as the constructed fusions

We next verified the expression of the constructed RNA and protein fusions. As shown by the Western Blotting Analysis in Figure 2.2A, successful expression of the four proteins (two fusion proteins, CsrA-NYFP and MS2-CFYP, and two non-interacting controls, NYFP and CYFP) was achieved from the constructed plasmids. Figure 2.2A confirms that the protein fusions are expressed at their predicted molecular weights (27.8 kDa for CsrA-NYFP, 25.3 kDa for MS2-CYFP, 21.0 kDA for NYFP, and 12.6 kDa for CYFP). In addition, we were able to detect stable dimer formation from CsrA-NYFP and MS2-CYFP at higher protein loading concentrations (Figure 2.2C). This observation is consistent with CsrA and MS2 being functional, since both CsrA and MS2 are known to form dimers (reviewed (83)). Also, an additional band appears when CsrA-NYFP and MS2-CYFP are co-expressed (as indicated by the dashed arrow in lane 3 of Figure 2.2C), suggesting that YFP is capable of refolding in the context of the fusion proteins. It should be noted that the electrophoresis of these proteins was conducted under denaturing conditions, but we suspect that the mercaptoethanol used to denature the proteins was not active enough to completely disrupt dimer formation.

In addition to testing the protein fusions, we needed to confirm that the RNA fusion was being expressed and was functional. Given that the CsrB fusion has additional nucleotides from the 2MS2bd or the 6MS2bd that have their own unique secondary structure, the extra nucleotides could interfere with the native structure of CsrB and its ability to bind CsrA *in vivo.* Functional CsrB was confirmed in our

CsrB-2MS2bd fusion by visual observation of a glycogen accumulation phenotype previously reported in the literature when CsrB is overexpressed (17, 28, 84). Figure 2.2B illustrates the effect of CsrB overexpression causing greater glycogen accumulation as visualized by the darker iodine staining in the cell pellet of the cells expressing CsrB (D) and CsrB-2MS2bd RNA (C) than in the cells expressing the 2MS2bd RNA species (A and B).

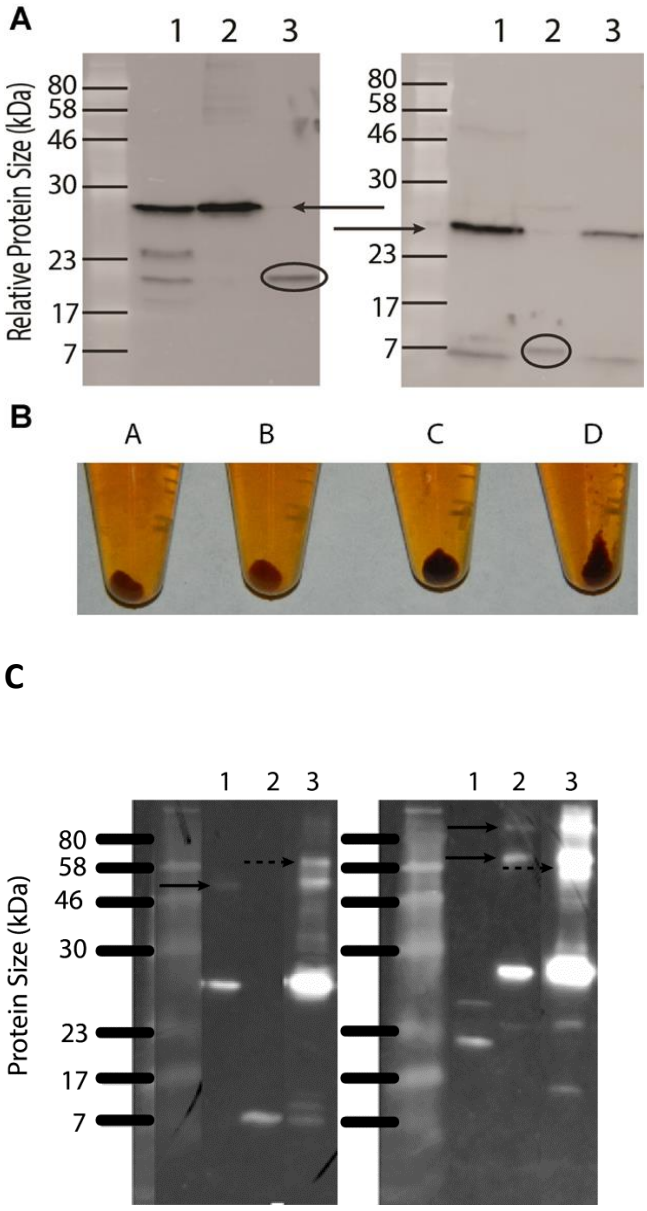


Figure 2.2: Evidence of protein and RNA expression

A. Western Blot Analysis: Left panel - Anti-FLAG Western Blotting analysis using total protein extracts from cells expressing the CsrA-NYFP and MS2-CYFP proteins (lane 1), the CsrA-NYFP and CYFP proteins (lane 2), and the NYFP and MS2-CYFP proteins (lane 3). Protein size markers are given on the left. Note that only the CsrA-NYFP and NYFP fragments are tagged with the FLAG epitope. The NYFP protein is 21.0 kDA and is

circled in lane 3. The CsrA-NYFP protein is 27.8 kDA and is marked by an arrow. Lower bands are likely degraded proteins. Right Panel - Anti-GFP Western Blotting analysis using total protein extracts from cells expressing the CsrA-NYFP and MS2-CYFP proteins (lane 1), the CsrA-NYFP and CYFP proteins (lane 2), and the MS2-CYFP and NYFP proteins (lane 3). Protein size markers are to the left. The protein MS2-CYFP is 25.3 kDA and is marked by an arrow. CYFP is 12.6 kDA and is circled. The smaller bands are likely degraded proteins.

B. Observation of CsrB phenotype in RNA fusions: These are cell pellets that have been stained with iodine to reveal glycogen accumulation. All cells were grown with the listed plasmid until stationary phase. A) pCsrA-NYFP + 2MS2bd expressed the CsrA-NYFP protein and the 2MS2bd RNA species. B) pNYFP + 2MS2bd expressed the NYFP protein and the 2MS2bd RNA species. C) pNYFP+CsrB-2MS2bd expresses the NYFP protein and the CsrB-2MS2bd RNA species. D) pUC19-CsrB expresses the CsrB RNA. The iodine staining is visually darker for tube C and D, indicating that glycogen accumulated as a result of expression of functional CsrB regulation.

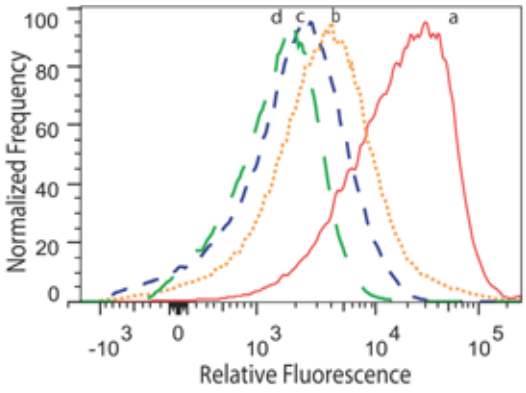
C. Western blot visualization of protein dimers: Left panel – anti-GFP blot for detection of CYFP proteins. Lane 1 contains the MS2-CYFP protein only; a solid arrow indicates a second band roughly double the size of MS2-CYFP indicating a properly folded MS2 dimer. Lane 2 is only the CYFP protein and shows no alternate bands. Lane 3 is protein extract from cells producing MS2-CYFP and CsrA-NYFP. A dashed arrow indicates the expression of a protein product larger than the MS2 dimer that likely corresponds to refolded YFP protein. Right panel – anti-FLAG blot for detection of NYFP protein. Lane 1 contains the NYFP protein. Lane 2 contains the CsrA-NYFP protein. Solid arrows indicate higher order products that correlate to proper CsrA dimerization. Lane 3 contains the proteins CsrA-NYFP and MS2-CYFP. A dashed line indicates the new protein band that is likely refolded YFP.

In addition to glycogen accumulation, we consistently observed cell adhesion to the walls of the culture vessel when CsrB-6MS2bd, CsrB-2MS2bd, or unaltered CsrB RNA was expressed, but no adhesion occurred in the controls expressing 6MS2bd or 2MS2bd RNA. The formation of biofilm is a known phenotype of *E. coli* overexpressing CsrB (85). This evidence confirmed that the 3' MS2bd tag did not completely interfere with the regulatory function of CsrB.

2.2.3 The CsrB-CsrA interactions enhance total fluorescence from fluorescence complementation

After confirming functional expression of all constructed fusions *in vivo*, we tested YFP refolding (as detected by fluorescence) under conditions that allowed or disallowed CsrB-CsrA interactions. To ensure that a fluorescent signal was only obtained in the presence of both CsrA and CsrB interacting components, various isogeneic control systems were constructed where one or more of the fusions lacked CsrA, CsrB, or the MS2 protein. For these experiments, combinations of the two plasmids carrying various forms of the genetic fusions were co-transformed into a mutant strain of *E. coli*, carrying a genomic knockout of the *csrB* gene (*ΔcsrB*::camR (25), a gift from Dr. Tony Romeo). It should be noted that CsrC, a complementary sRNA to CsrB, was still expressed in the cell and could potentially interfere with CsrB. However, CsrC has a much lower affinity for CsrA than CsrB (18) and is expressed at lower concentrations than CsrB (86), so the CsrB fusion RNA should preferentially bind RNA before CsrC.

Cells were grown to stationary phase and the fluorescence of the cell population was measured using flow cytometry. As shown in Figure 2.3, fluorescence is enhanced only when the full system was expressed (curve a). Only modest shifts in fluorescence are observed in any of the controls where the sequence for CsrA, CsrB, or MS2 is absent.



RNAs and Proteins present	a	b	c	d
CsrA-NYFP	+	+	-	+
NYFP	-	-	+	-
MS2-CYFP	+	+	+	-
CYFP	-	-	-	+
CsrB-2MS2bd	+	-	+	+
2MS2bd	-	+	-	-
Fluorescent Measurements	a	b	c	d
Geometric Mean	16765	3517	2052	1301
Mean	26111	4993	3162	1819
Median	19452	3768	2326	1547
Mode	30597	4193	2849	2091
Standard Deviation	200225	14265	2053	2081
Count	60012	104974	83536	65737
T(X)	12900	5400	1820	-

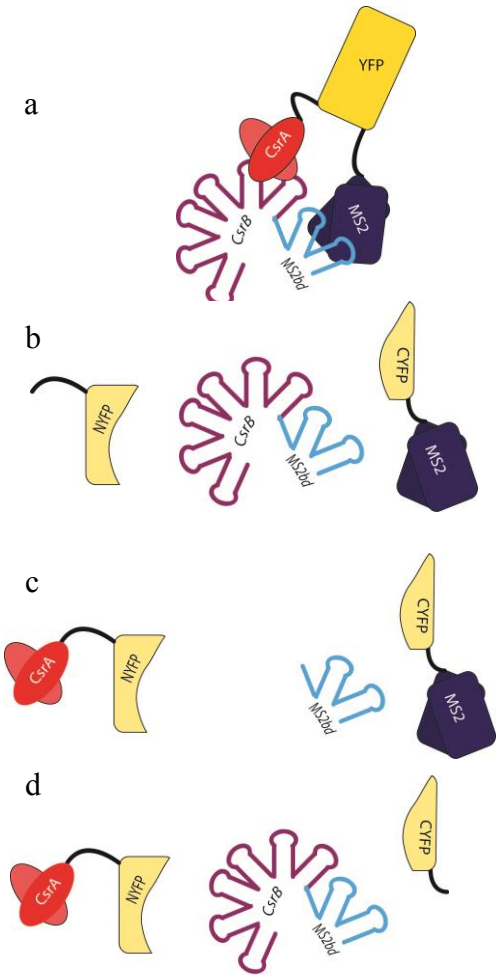


Figure 2.3 Fluorescence complementation occurs in the presence of the three fusion molecules

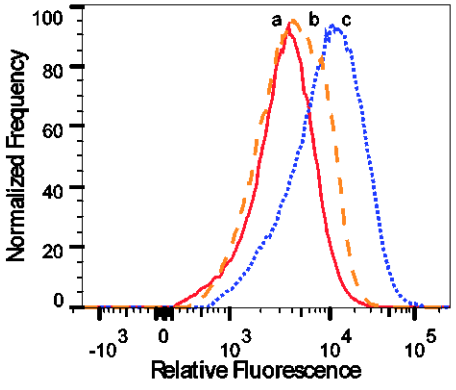
Each curve represents the population fluorescence for one of the system controls. a) Red solid line: Full system (CsrA-NYFP, CsrB-2MS2bd, and MS2-CYFP). b) Orange dotted line: No CsrB sequence control (CsrA-NYFP, 2MS2bd, and MS2-CYFP). c) Blue dashed line: No CsrA sequence control (NYFP, CsrB-2MS2bd, and MS2-CYFP). d) Green long-dashed line: No MS2 sequence control (CsrA-NYFP, CsrB-2MS2bd, and CYFP). The chart below summarizes the fusions present in each system and gives the numerical value of relevant fluorescence statistics. These experiments were done in a *ΔcsrB*::*camR* strain of *E. coli*.

In order to quantitatively compare changes to fluorescence, we analyzed the difference in populations with the probability binning statistical method that is available in the FlowJo software (87). Probability binning is a modified chi-squared statistic that compares two univariate distributions. Probability binning is a preferable method to compare populations of exponentially distributed data as the method is not sensitive to the large variation in the standard deviation that occurs in more fluorescent populations. In this method, the control population distribution is divided into bins so that each bin has the same number of events. These bins are then compared to sample distributions to calculate a chi-squared value, T(χ), which represents the number of standard deviations that the observed chi-squared value differs from the minimal meaningful value. This value is analogous to a t-score, and a T(χ) value greater than four predicts that two populations are different with 99% confidence. More importantly, this value can also be used to rank samples according to their similarity to a control population. Using this method to compare histograms, the T(χ) value has been calculated using the lowest fluorescent system, the no MS2 control (curve d), as the control population; these values are given in Figure 2.3. As expected, this analysis shows a significant shift in the fluorescence distribution when the fluorescence recombination is directed by the CsrB-CsrA interaction.

The importance of using the CsrB knockout was tested by measuring the ability of the CsrB-CsrA interactions to direct fluorescence complementation in the wild type *E. coli* (MG1655), where native CsrB is present. As shown in Figure 2.4A, a significant shift in fluorescence is not detected with expression of the full system in

the wild type strain (in contrast to when the full system is expressed in the *ΔcsrB*::*camR* mutant strain, as seen in Figure 2.3). An explanation for this observation is that native CsrB outcompeted the designed CsrB fusion for CsrA-NYFP. This explanation is supported by the comparative Northern Blotting analysis of CsrB expression (Figure 2.4B). In this blot, the CsrB fusion was expressed in the wild type strain and the *ΔcsrB*::*camR* mutant strain. The blot shows the native CsrB is much more prevalent than the CsrB fusion in the wt cells.

A



RNAs and Proteins present	a	b	c
CsrA-NYFP	+	+	-
NYFP	-	-	+
MS2-CYFP	+	+	+
CYFP	-	-	-
CsrB-2MS2bd	+	-	+
2MS2bd	-	+	-
Fluorescent Measurements	a	b	c
Geometric Mean	3787	4174	10422
Mean	4670	5687	15524
Median	3910	4325	11507
Mode	3926	4481	11534
Standard Deviation	3437	5245	14554
Count	72891	75327	68780

B

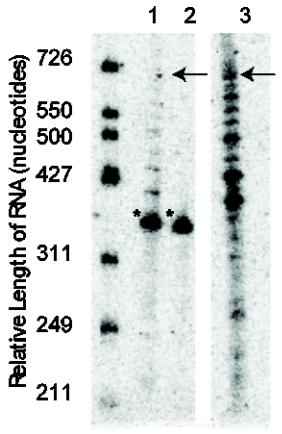


Figure 2.4: TriFC is not observable in wt MG1655 cells

A. Histograms and fluorescence statistics of TriFC system in MG1655. When all necessary proteins and RNAs for fluorescence complementation (population a, solid red) are present in MG1655, the observed fluorescence is lower than the control lacking the CsrB-2MS2bd RNA (population b, dashed orange line) or the control lacking the full protein fusion CsrA-NYFP (population c, dotted blue line)

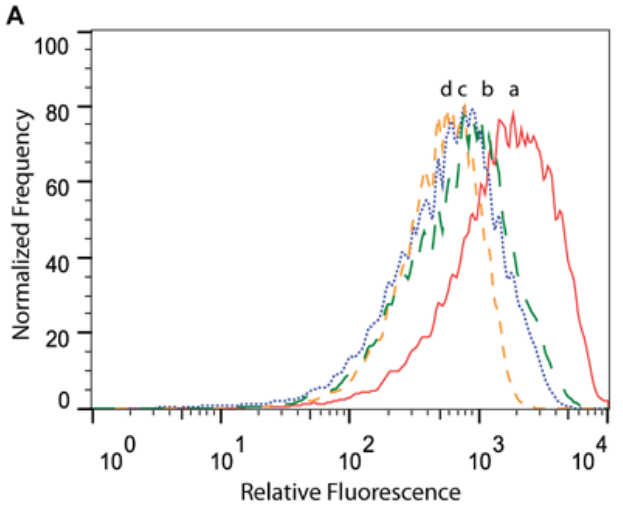
B. Northern blot of expression of native and fusion CsrB. Lane 1 - MG1655 cells expressing a CsrB-6MS2bd variant of the fusion RNA. Lane 2 - RNA from MG1655 expressing a MS2 binding domain. Lane 3 - *ΔcsrB*::*camR* mutant strain expressing CsrB-6MS2bd variant of fusion. The star symbol marks native CsrB RNA, and the arrow marks CsrB-6MS2bd RNA.

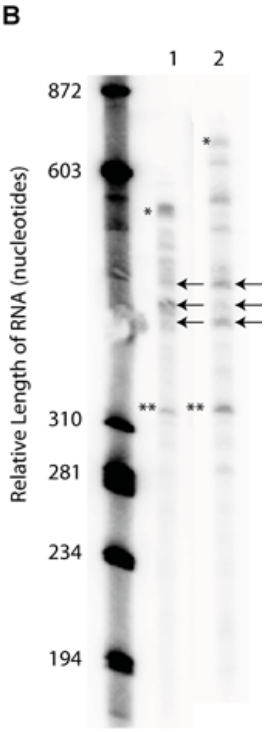
2.2.4 The sequence of the RNA binding domain affects the ability of the CsrB fusion to enhance fluorescence

We next attempted to enhance the fluorescence signal by increasing the number of MS2 binding sites in the original RNA molecular design. Our rationale was that a higher number of binding sites would recruit more MS2-CYFP fragments to the RNA fusion and provide a better chance for fluorescence complementation to occur. This rationale is supported by in the literature where mammalian mRNAs have been tagged with up to 24 MS2bd repeats for effective visualization (88). Likewise, higher numbers of the hairpin repeats have been reportedly used to purify RNAs and RNPs (81).

For our system, another RNA construct was made expressing a set of six MS2 hairpins (6MS2bd) and was compared to the 2MS2bd RNA. Surprisingly, the 6MS2bd construct was unable to produce fluorescence complementation as a fusion with CsrB (Figure 2.5A).

In contrast, fluorescence complementation was only observed with the 2MS2bd system (discussed above). For comparison purposes, the T(χ) value of CsrB-6MS2bd and the 6MS2bd fluorescent populations (curves b and c) were calculated using the 2MS2bd population (curve d) as a control. These T(χ) values are much closer to each other than to the CsrB-2MS2bd population (curve a) (it should be noted that fluorescence values were measured on a different model cytometer than the curves presented in Figure 2.3).





RNAs and Proteins present	a	b	c	d
CsrA-NYFP	+	+	+	+
MS2-CYFP	+	+	+	+
CsrB-2MS2bd	+	-	-	-
2MS2bd	-	-	-	+
CsrB-6MS2bd	-	+	-	-
6MS2bd	-	-	+	-
Fluorescent Measurements	a	b	c	d
Geometric Mean	1537	522	676	479
Mean	2109	803	986	599
Median	1685	626	757	528
Mode	2371	866	1037	562
Standard Deviation	1597	680	874	373
Count	23557	23679	24714	24535
T(X)	2700	400	700	-

Figure 2.5: Comparison of two variations of the MS2bd RNA sequence

A. Fluorescence of cells expressing the 2MS2bd RNA fusion and the 6MS2bd RNA fusions in *ΔcsrB*::*camR*: a) Red solid line: Full system with CsrB-2MS2bd. Cells express CsrA-NYFP, MS2-CYFP, and CsrB-2MS2bd fusions. b) Green long-dashed line: Full system with CsrB-6MS2bd. Cells express CsrA-NYFP, MS2-CYFP, and CsrB-6MS2bd fusions. c) Blue dotted line: 6MS2bd control. Cells express CsrA-NYFP, MS2-CYFP, and 6MS2bd fusions. d) Orange dashed lines: 2MS2bd control. Cells express CsrA-NYFP, MS2-CYFP, and 2MS2bd fusions. The full system with CsrB-2MS2bd (line a) shows a positive shift in fluorescence, while the full system with CsrB-6MS2bd (line b) is roughly equivalent to the no CsrB fusion RNA controls for the 2MS2bd and the 6MS2bd (lines c and d). The fluorescence values were taken using a FACSCalibur (BD Biosciences).

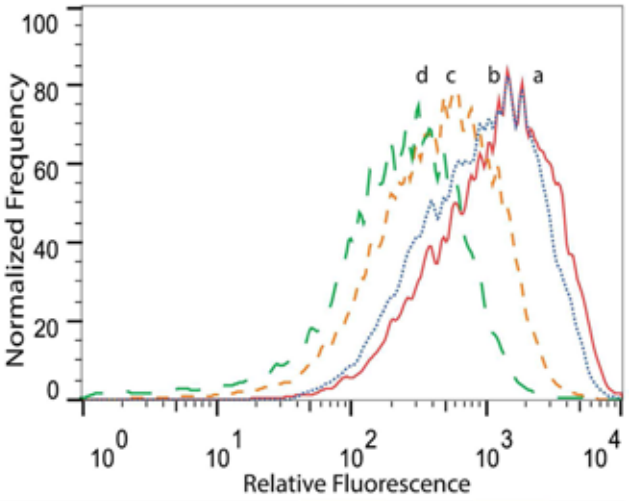
B. Northern Blot analysis of the CsrB-2MS2bd (lane 1) and CsrB-6MS2bd (lane 2) RNA fusions in *ΔcsrB*::*camR*. Samples from total RNA extracts

(normalized to the same amount of RNA in each case) were probed for CsrB (see Methods). The * denotes the expected full length of the CsrB-2MS2bd or CsrB-6MS2bd RNA transcript (580 or 750 nucleotides, respectively). Both transcripts appear to have transcriptional termination very near the transition between the CsrB and the MS2bd at roughly 330 nucleotides (marked with a **). In addition, both transcripts appear to have transcriptional terminators in the same locations for the next three bands (marked with arrows).

To further understand why the 6MS2bd failed to enhance fluorescence, the 6MS2bd and 2MS2bd constructs were analyzed by Northern Blotting analysis using a CsrB probe. As shown in Figure 2.5B, both fusions show premature transcriptional termination. It is expected that the full length CsrB-2MS2bd (lane 1) should be 580 nucleotides and CsrB-6MS2bd (lane 2) should be 750 nucleotides. These shortened transcripts are likely the result of transcriptional termination, as we only observed these sharp bands for RNA lengths corresponding to the beginning of the MS2bd of the RNA fusion (at lengths longer than 320 nucleotides). While the banding patterns of both RNA transcripts are similar, the CsrB-2MS2bd shows a significant change in banding patterns near the end of its sequence when compared to CsrB-6MS2bd. An interpretation of these observations is that CsrB-2MS2bd has a more beneficial conformation than the CsrB-6MS2bd for binding either the MS2-CYFP or CsrA-NYFP fusions to direct fluorescence complementation. It is possible that a higher number of MS2bd repeats may not be the optimal tag for a regulatory RNA that is expected to require high structural flexibility, rapid configuration changes, and the ability to sequester a high number of proteins (as in the case of the CsrB). In addition, it appears that the tagging of RNAs with a high number of repeated genetic palindromes (such as the multiple MS2bd sequences and the repeated cloning sites) leads to multiple conformations that are transcriptionally unstable or less likely to bind the MS2 target. Overall, these results are counterintuitive to our expectations that more binding sites would enhance the fluorescence values.

2.2.5 TriFC can detect a dynamic range of RNA-protein interactions

To investigate the sensitivity of the fluorescence response of this system to the CsrB-CsrA interaction, several CsrA mutants were constructed with altered affinity to the target CsrB RNA. The mutants of CsrA that were examined were V40A, V42A, and R44A (where the K_d for the association exhibits increases of 60-fold, 120-fold, and 150-fold, respectively, compared to wild type); these mutations were designed based on available *in vitro* biochemical characterization of this system (74). It is worth noting that these K_d values can only be used as qualitative comparative trends between the mutants since they were previously determined *in vitro* by gel-shift assays between CsrA and a single SELEX-optimized RNA hairpin motif. As shown in Figure 2.6, these mutants displayed a pattern of fluorescence intensity *in vivo* that reflects observations from the published *in vitro* experiments.



System	a	b	c	d
Mutation	wt	V40A	V42A	R44A
Fluorescent Measurements	a	b	c	d
Geometric Mean	830	836	386	189
Mean	1672	1230	570	288
Median	948	931	423	231
Mode	1433	1433	562	229
Standard Deviation	1081	1060	552	285
Count	20782	9583	17228	20614
T(X)	2100	1100	600	-

Figure 2.6: Relative fluorescence values from mutants of CsrA expressed in the CsrA-NYFP fusion

Fluorescence distribution of TriFC system with mutations to the CsrA portion of the CsrA-NYFP fusion protein a) Red solid line: wt CsrA, b) Blue dotted line: V40A, c) Orange dashed line: V42A, and d) Green long-dashed line: R44A.

The least detrimental mutation to RNA binding affinity, V40A, yields almost no change in fluorescence. The next most detrimental mutation, V42A, modestly reduced fluorescence, and the most detrimental mutation, R44A, significantly reduced fluorescence to levels similar to the no interaction controls (data not shown). Using the R44A distribution as a reference, the T(χ) value reflects the increasing fluorescence of the population distributions (it should be noted that the T(χ) value depends on the size of the population as well, so the T(χ) of V40A is not directly comparable to the wt CsrA). Collectively, these observations are significant as they quantitatively correlate RNA-protein binding affinities to fluorescence readouts from this system. Importantly, this dynamic detection range confirms this system's ability to capture a range of interaction affinities between CsrB and CsrA in *E. coli*. This type of quantitative nature to a three partite complementation assay had not yet been demonstrated in the context of regulatory sRNA-protein interactions in bacterial systems. Although the *in vitro* data predicts that all mutants would exhibit fluorescence changes relative to the wild type CsrA, it is possible that the presence of CsrA heterodimers in our study (where one CsrA is natively-encoded and the other is a mutant plasmid-encoded NYFP fusion) mitigates the negative effect of the mutation on CsrB-CsrA binding affinity (32). Additionally, the mutants were characterized using single RNA hairpins; CsrB, with its multiple CsrA binding sites, may behave differently with these mutants.

2.3 DISCUSSION

The adaptations to the TriFC method developed in this work present a valid technique to study RNA-protein interactions *in vivo* in bacteria. Previous works using this system focused on using TriFC as a method to study RNA-protein localization (64), RNA visualization (77), or RNA dynamics (89). We have demonstrated the potential of this system to quantitatively assay regulatory sRNA-protein interactions in the context of their native pathways through fluorescence complementation. The ability of this system to capture these regulatory interactions in the context of the Csr network is highly promising, given the unique challenges presented by the complex geometry of its interactions. First, in this system, fluorescence complementation depends on the ability of a tagged CsrB RNA (CsrB-*2*MS2bd) to bind its target, CsrA, while the CsrA is fused to a foreign protein (NYFP). Secondly, the fusion interaction is further complicated by the observation that the number (and presumably configuration) of the MS2bd can have significant impact on the RNA structure. Lastly, CsrB is known to bind multiple copies of CsrA, potentially making it more challenging for the CsrA-NYFP chimera to bind the MS2 tagged CsrB in a position favorable for YFP refolding. Even within this complex arrangement, directed fluorescence complementation can detect sRNA-protein interactions, and it should be possible to adapt this system for other RNA-protein pairs. In Chapter 3, this method is adapted to detect mRNA-CsrA interactions.

The ability to discriminate between various CsrA mutants presents the possibility of probing mutant libraries of RNA-protein interactions *in vivo*. A natural

advantage of TriFC is its potential for high-throughput applications given the relative ease of using cytometry to screen large libraries of mutants based on their fluorescent values. It becomes possible to envision the rapid identification of mutations that affect the binding affinity. This evaluation would be instructive to understand the critical residues that dominate intermolecular interactions of regulatory RNAs and their protein targets *in vivo*. This technique would serve to compare binding affinities of RNA variants generated by *in vitro* SELEX techniques to their binding capabilities observed *in vivo*. Lastly, this tool presents the potential to quantitate how interactions among cellular regulators change from outside stimuli triggering global adaptation responses.

2.4 METHODS

2.4.1 Construction of the MS2-CYFP containing plasmid

The plasmid pMS2-CYFP encodes for the MS2 coat protein fused to the C-terminal fragment of YFP (aa 155-238) by the linker developed by Hu et al. 2002 (80). The MS2 coat protein contains the deletions that prevent capsid formation that occur in the wild type coat protein (79). The plasmid pMS2-CYFP was adapted from the plasmid pDMSD-Y2 previously used to detect fluorescence between the proteins DmsA and DmsD (90), which originates from the vector pKNT25 (Euromedex). The MS2-CYFP protein sequence is inserted in between the HindIII and ClaI recognition sites of the vector backbone pKNT25. An alternative plasmid was constructed with only the CYFP sequence in this location to produce a no MS2 control. The MS2 coat protein and the plasmid pDMSD-Y2 were gifts from Dr. Matt DeLisa.

2.4.2 Construction of the CsrA-NYFP protein encoding plasmid

The plasmid pCsrA-NYFP encodes for the CsrA polypeptide (17) fused to the N-terminal domain (amino acids 1-154) of YFP by a fifteen amino acid linker consisting of 3 repeats of GGGGS. This sequence was constructed with the plasmid pDMSA-Y1, which was a gift from Dr. Matt DeLisa (90). The NYFP protein contains a C-terminal FLAG tag in order to be detected by Western Blot. Standard cloning techniques were used to create this sequence. An alternative plasmid was constructed to create a no CsrA control, in which the linker-NYFP sequence was inserted into the pDMSA-Y1 plasmid to create the plasmid pNYFP.

2.4.3 Construction of the RNA encoding fusion

The RNA fusion construct was synthesized to express CsrB (30) and a sequence containing the MS2 binding domain (MS2bd) using the pUC19 vector. The CsrB sequence was a gift from Dr. Tony Romeo. The MS2 binding domain includes the high affinity MS2 coat protein binding domain (78, 82). This sequence was a gift from Dr. Marlene Belfort. There were two variants of the MS2bd used in this study. One variation of the MS2bd contains six of the high affinity hairpins (6MS2bd) and is a series of self-ligations of the sequence proposed by Bardwell and Wickens. The other variation contains only two of the hairpins (2MS2bd). The 2MS2bd was constructed after the 6MS2bd, and the process is described in the next section.

The 6MS2bd was inserted into the CsrB RNA at position 320 of the CsrB sequence. This nucleotide was chosen to insert the MS2bd because it is upstream of the stem loop predicted to be part of the transcriptional terminator (30). In order to

construct the RNA fusion in pUC19, the *lac* promoter was manipulated at the M13-rev sequence (5'-CAGGAAACAGCTATGACCATG) to include the NdeI cut site at the beginning of the pUC19 translation initiation site (5'-CAGGAAACAGCATATGACCATG) using the Quick Change II Site Directed Mutagenesis kit (Agilent). The 6MS2bd was inserted into the pUC19 vector at the newly formed NdeI site and the AatII restriction site. The stop transcription site of CsrB was added to the 3' end of the MS2bd. This created the 6MS2bd RNA sequence that would be used as a no CsrB control. The 5' section of CsrB was then added to the 5' end of the 6MS2bd using the restriction sites of NdeI and KasI. (The KasI sites were part of the 6MS2bd.) This created the CsrB-6MS2bd RNA fusion.

2.4.4 Construction of the two-promoter RNA and protein encoding plasmid and CsrA mutants

To simplify expression of the three interacting elements (NYFP fusion, CYFP fusion, and RNA fusion), the CsrA-NYFP and CsrB-MS2bd fusions were combined in the same plasmid. This system was constructed by amplifying the CsrB-6MS2bd RNA fusion up stream of its *lac* promoter with the primers: 5'-CCTGACGTCGCGAGGAAGCGG and 5'-TTAGACGTCAATAAAAAAAGGGAGCAC. The amplified genetic material was digested with AatII and ligated into plasmid pCsrA-NYFP. This created the plasmid pCsrA-NYFP+CsrB-6MS2bd. This plasmid was modified with several mutations to include additional restriction sites to aid in the insertion of additional genetic sequences to the CsrA or the CsrB positions on the plasmids. A second version of this

plasmid was constructed to reduce the size of the RNA fusion by reducing the 6MS2bd from six repeats of the MS2 hairpin (6MS2bd) to two repeats of the hairpin (2MS2bd). This was done by digesting the plasmid with KpnI and self-ligating the plasmid to create the plasmid pCsrA-NYFP+CsrB-2MS2bd.

Several mutant varieties of CsrA were introduced into the CsrA-NYFP protein in the plasmid pCsrA-NYFP+CsrB-2MS2bd. The mutations to CsrA were V40A, V42A, and R44A, as these sites have been previously identified as important to CsrA function (74). These mutations were constructed using the protocols of the Quick Change II Site-Directed Mutagenesis kit (Agilent).

Negative controls for this system were created by removing the CsrA, CsrB, MS2 genetic elements from these plasmids using standard cloning techniques. These plasmids are named for the proteins and RNAs that they express. For a list of plasmids created by this study see Table 2.1: Plasmids and Fusion Expression.

Table 2.1: List of plasmids with protein and RNA fusion products used in this study

Plasmid	Protein Fusion	RNA Fusion
pCsrA-NYFP+CsrB-*2*MS2bd	CsrA-NYFP	CsrB-*2*MS2bd
pCsrA-NYFP+CsrB-*6*MS2bd	CsrA-NYFP	CsrB-*6*MS2bd
pCsrA-NYFP+*2*MS2bd	CsrA-NYFP	*2*MS2bd
pCsrA-NYFP+*6*MS2bd	CsrA-NYFP	*6*MS2bd
pNYFP+CsrB-*2*MS2bd	NYFP	CsrB-*2*MS2bd
pNYFP+CsrB-*6*MS2bd	NYFP	CsrB-*6*MS2bd
pMS2-CFYP	MS2-CFYP	-
pCYFP	CYFP	-

2.4.5 Cell culture and expression of genetic fusions

The strain RG1-B MG1655 (25), a gift from Dr. Tony Romeo, is a genomic knock out of CsrB and was selected for expression of all vectors in this study. The RG1-B MG1655 strain was transformed with plasmids sequentially using $CaCl_2$ transformation protocols. The first plasmid transformed into cells was the pMS2-CYFP or the pCYFP plasmid, respectively, and was followed by the CsrA-NYFP+CsrB-2MS2bd vector, or one of its variants. After cells were transformed with both plasmids, individual colonies were inoculated into shaker flasks containing 40 mL of Luria Broth (LB) augmented with the antibiotics kanamycin and ampicillin (to reflect the two plasmids carried by the cell). The cultures were grown at 25ºC with agitation. All genetic fusions were under the control of the *lac* promoter. However, as the *lac* promoter is slightly leaky, the addition of inducer was not used to control expression. At least three different cell populations from each plasmid combination were grown for analysis by flow cytometry.

2.4.6 Collection and analysis of fluorescence data

Cells were analyzed by flow cytometry using the FACSAria III flow cytometer and the FACSCalibur (BD Biosciences). Cells were prepared for cytometry by diluting the cells into phosphate buffered saline (PBS: 137 mM sodium chloride, 2.7 mM potassium chloride, and 10 mM phosphate buffer, pH 7.4-7.6) to a concentration on the order of 10^7 cells/mL. The cells were excited with the 488 nm laser found in the flow cytometer, and the cell population was determined from the forward scatter and side scatter distributions reported by the cytometer. Data was collected for at least

25,000-50,000 active cells. The high number of events collected ensures enough events to assume that the population distribution would be unaffected by rare events. For this study, all cells are analyzed during the population's stationary phase.

Analysis of the cytometry data was done using the FlowJo software (Tree Star, Inc.). Examination of the fluorescence values showed that the forward scatter and side scatter of the true cell population overlapped with non-fluorescent background noise. In order to separate the fluorescent cellular population and non-fluorescent background signal, the events were plotted on a dot plot of the fluorescence covered by the FITC vs. PE band pass sensors of the cytometer. These two filters were selected because they cover the range of emission values of YFP fluorescence and are able to differentiate fluorescence due to YFP and the background level. From this comparison, it was determined that fluorescence from YFP could be differentiated from the background using the difference in the FITC-A fluorescence of the two populations. All events that were above the FITC-A fluorescence of the background noise were considered to be living cells and were compared using the PE-A reading to calculate the mean, geometric mean, median, mode, and standard deviation of fluorescence. These calculations were made by the FlowJo software. The histograms of fluorescence were normalized and presented using suggested parameters by FlowJo, which uses the mode of a reference population to calculate the scaling factor between populations.

2.4.7 Protein analysis - extraction and Western Blotting

Soluble proteins were extracted from 10 mL of saturated cell culture that was grown as described in the Cell Culture and Expression of Genetic Fusions section (Section 2.4.5). The cells were collected by centrifugation and re-suspended in 1 mL of PBS in a 1.5 mL conical centrifuge tube. Cells were lysed by sonication with the Qsonica 55 probe (Qsonica, LLC). Cells were kept on ice and sonicated for 60 seconds, maintaining between 10-15 W of output energy. The soluble portion of the cell lysate was quantified for total protein content using a Bradford assay (ThermoScientific) relative to standards of bovine serum albumin (New England Bioloabs). Electrophoresis was conducted using the Miniprotean (BioRad) electrophoresis set up. The wells were loaded with 5 µg of soluble protein diluted into 40 µL with an appropriate amount of denaturing SDS sample loading buffer and heated to 95ºC for 5 min. The gel was an SDS-PAGE gel prepared with the Laemmli setup using a 12% resolving gel and a 6% stacking gel. Samples were run with the ColorPlus™ Prestained Protein Marker, Broad Range (NEB) as a standard. Gels were transferred to a nitrocellulose membrane using the TranBlot SD Semi-Dry Electrophoretic Transfer Cell (BioRad). Gels were prepared by soaking in Bjerrum and Schafer-Nielson transfer buffer with SDS for 10 min. The transfer was decided to be sufficient when the colored protein marker at 46 kDa was no longer visible in the gel. Blotting was done with standard procedures at room temperature on a rocking platform. The initial blocking step was conducted with 5% dry milk in Tris-buffered saline solution (TBS). The protein fusions of NYFP and CYFP were detected using

the Anti-FLAG M2 antibody (Agilent) and the Anti-GFP (Roche) as a primary antibody, respectively. Both antibodies were diluted to the manufacture specifications in TBS and 1% w/v dry milk. The secondary antibody used in both cases was the Anti-Mouse IgG (H+L), HRP Conjugate (Promega), diluted to the manufacturers specifications in TBS and 2% (w/v) dry milk. All images were developed using the Immun-Star™ HRP Chemiluminescent kit (BioRad) and the ChemiDoc™ MP Imaging System (BioRad).

2.4.8 RNA analysis: RNA extraction and Northern Blotting

RNA was extracted from cell pellets collected from 2 mL of saturated liquid culture using the TRIzol® (Life Technologies) RNA extraction reagent. Cells were lysed in the TRIzol reagent using silica beads and a bead beater to agitate cells for 100 seconds. The RNA extraction followed the TRIzol instructions. RNA was quantified using absorbance at 260 nm detected by the NanoDrop™ Spectrophotometer (Thermo Scientific). Electrophoresis was conducted using the Protean II™ (BioRad) electrophoresis array and using 5% denaturing Urea-PAGE gels. 10 µg of total RNA was loaded into the gel. Radio labelled PhiX174 DNA/HaeIII Markers (Promega) were used as molecular weight standards. Radio labelling was conducted using gamma P32 ATP (Perkin Elmer) and T4 polynucleotide kinase (NEB). Transfer was performed using a Hybond-N+ (GE Healthcare Life Sciences) nylon membrane and the Trans-Blot™ (BioRad) transfer system. The RNA was fixed to the membrane through UV crosslinking using the HL-2000 HybriLinker™ (UVP) system using the UV settings of 1200 µJ/cm2.

RNA membranes were blotted with the DNA probe "CsrB Probe 4". The sequence of CsrB Probe 4 is 5'-CCGAATCATCCTGACCGG-3' and binds the CsrB RNA between nucleotides 152-170 of CsrB. All probes were radiolabeled with P32 using gamma P32 ATP (Perkin Elmer) and T4 polynucleotide kinase (NEB). Blotting was carried out by incubating crosslinked membranes with 20 pmol of radiolabeled DNA with 15 mL of hybridization buffer (Perfect Hyb™ Plus hybridization buffer (Sigma)) at 42ºC with rolling agitation for at least 1 hour. The membranes were then incubated with solutions of saline sodium citrate with 0.1% w/v SDS (1X SSC is 150 mM sodium chloride, 15mM trisodium citrate, pH 7.0). The membrane was first incubated with 5X SSC for thirty minutes followed by two incubations thirty minute incubations with 1X SSC. The membranes were washed with distilled water until all SDS was removed. The images of the membrane were taken using a phosphor screen (GE Healthcare) and developed using a Typhoon™ biomolecular imager (GE Healthcare).

Chapter 3

Translational assay for characterizing CsrA regulation on mRNA transcripts with interaction validation using TriFC

3.1 INTRODUCTION

The regulatory protein CsrA has been previously described as a global regulator of translation. As such, a better understanding of the targets regulated by this system is of vital importance to understanding cellular regulation. This knowledge could also improve genetic models of regulatory networks, which are often limited to transcriptional models (91, 92). Previous attempts to characterize CsrA's "targetome" have focused on either the identification of mRNA-CsrA interactions using crosslinking and immunoprecipitation (CLIP) techniques or analyzed proteomic, transcriptomic, and metabolomic data of Csr modified systems (21). Although both techniques are powerful tools to characterize complex systems, they both fail to make definitive claims of regulatory interactions. CLIP is often criticized because the method identifies inconsequential reactions (50), and omics data is unable to identify direct interactions. Ideally, any attempt to characterize a regulatory system would showcase direct interactions and utilize the regulatory mechanism in order to demonstrate the capacity of the regulatory interaction.

We have developed a translational based assay to characterize the effects of the presence of CsrA on specific mRNA targets. In our assay, an mRNA target of interest is modified to produce a translational fusion with GFP to provide a fluorescence signal that correlates to the translational activity of the mRNA target. We utilized a plasmid-based expression system to control the expression of CsrA. Changes in the fluorescence level in relation to CsrA expression indicate a regulatory interaction occurring between the mRNA and CsrA. Furthermore, this system ameliorated many of the issues caused by indirect regulatory interaction by moving the expression system to a plasmid based

system. Approximately 250 mRNA targets were evaluated for their response to CsrA expression using this translational assay. These targets were largely identified from omics based studies of the Csr system that were done in collaboration with Steven Sowa (publication in review).

These regulatory interactions identified in the translational assay were further characterized by adapting the TriFC method presented in Chapter 2 to detect specific mRNA-protein interactions. Taken together, the characterization of the physical regulatory response and the evidence for direct mRNA-CsrA interactions provide strong evidence for the identification of 19 targets of CsrA previously undocumented in the literature (Table 3.4).

3.2 RESULTS

3.2.1 The 5' UTR translational assay

We augmented a translational assay to assess the effect CsrA has on the translational activity of an mRNA target. In this system, the mRNA reporter consisted of the 5' UTR of the mRNA of interest and 100 nucleotides of the coding sequence (CDS) attached in frame to a GFP reporter. The strength of translation from the targeted mRNA is related to the fluorescence intensity. It was decided that the translational reporter would test the 5'UTR and 100 nucleotides of the coding sequence based on the observation that most documented CsrA sites bind mRNA in the 5'UTR and near the translation initiation site to alter translational activity (41).

Targets were examined for their fluorescence intensity in the presence and absence of excess CsrA to determine the effect that CsrA had on translation. A schematic

of the experiment and some potential outcomes of the experiment are presented in Figure 3.1.

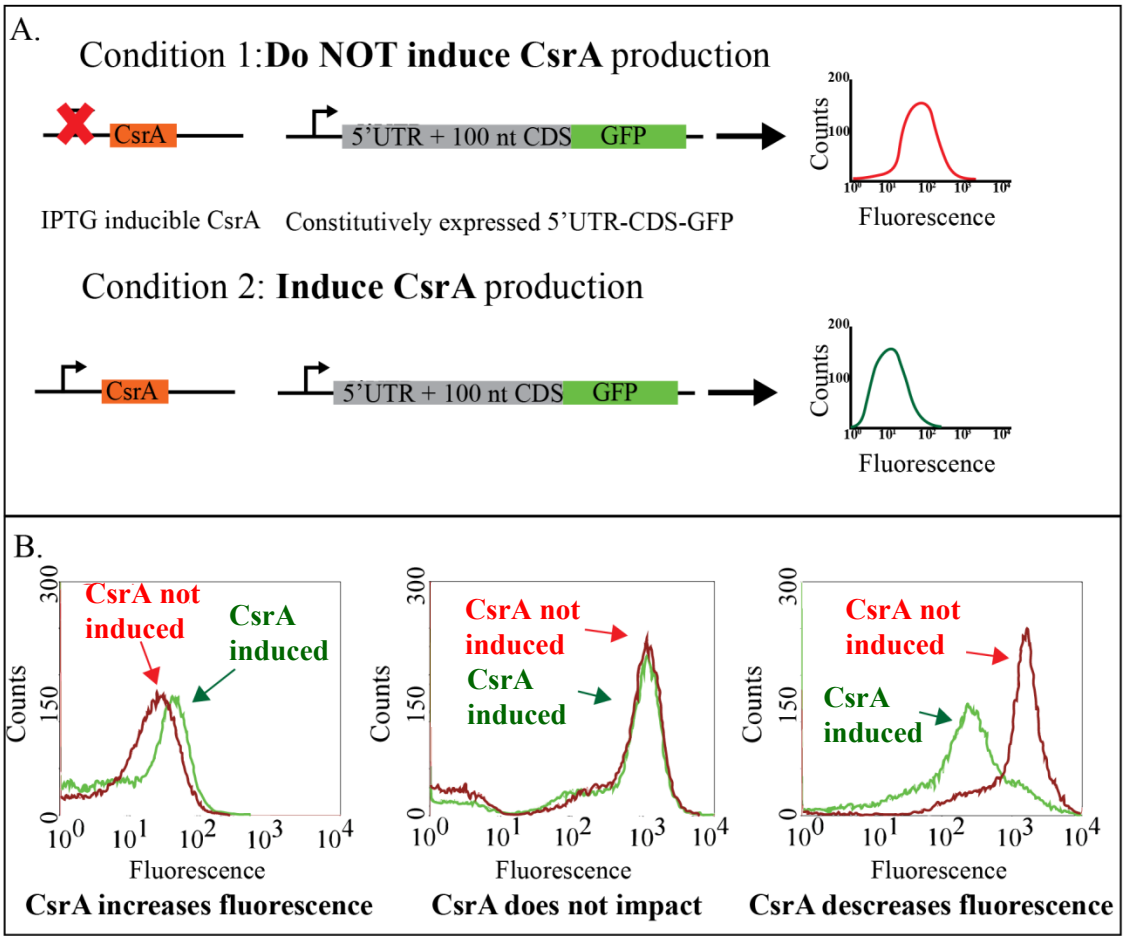


Figure 3.1: Schematic of 5'UTR translation assay with sample outcomes

A. Concept of the 5'UTR translation assay: The system is composed of two elements, the inducible CsrA expression system, and a 5'UTR and CDS of a gene of interest. In this system, the mRNA of interest is produced constitutively. The mRNA target of interest translates a section of its own protein fused in frame to a GFP signal that is easily detectable by flow cytometry. Evaluation of the effects of CsrA on the test sequence compares the fluorescence output when CsrA is present to when it is not present.

B. Sample data of the three observable outcomes of CsrA's relationship to target mRNA translation: Left panel, CsrA expression increases observable fluorescence of population. Middle panel, CsrA expression does not affect fluorescence of population. Right panel, CsrA expression reduces fluorescence of population.

The system adapts a translational assay from recent work analyzing the kinetics of the Csr system (93). In that work, the authors created a translational fusion of the 5'UTR and translational initiation site of *glgC* and GFP to understand how perturbations to the concentration of CsrA or CsrB affected the translation from *glgC*. In this work, the 5'UTR assay was developed from their plasmids pHL1756 and pHL600 and their modified *E. coli* strain, HL4142. All strains and plasmids are available from Addgene ® services.

The plasmid pHL1756 contains a constitutively active promoter that is capable of producing a transcript that contains the 5'UTR of a selected mRNA and the coding sequence for GFP. In this study, we modified the segment of the plasmid containing the 5'UTR sequence to investigate mRNA targets of interest. The plasmid pHL600 contains an inducible CsrA protein coding sequence that can be used to alter the concentration of CsrA. These plasmids were expressed in the strain HL4142, which contains genetic knockouts of the genes *pgaA, glgC, csrA, csrB, csrC,* and *csrD.* Removing these genes ensures that native expression of the Csr system doesn't affect the inducible, plasmid based components. Removing *pgaA* prevents biofilm development that can occur in a Csr deficient strain. It is important to note that *csrA* has been described as essential in the past, but recent work has shown that cells can survive a knockout of *csrA*, provided that the gene *glgC* is also removed (20).

3.2.2 Selection of mRNA targets

Potential mRNA targets were selected using two sources: 1) experimental data from an omics study of Csr mutants that included CsrA co-purification assays and 2)

sequence based predictions of CsrA binding. The experimental data is from a work that was done in collaboration with Steven Sowa (in review). The model predictions are based on the presence of the relative number of GGA based CsrA binding motifs present in the 5'UTR and the first 100 nucleotides of the coding sequence. Genes that had more than 3 such motifs were selected to be analyzed by the translational assay. This method of predicting CsrA targeted mRNAs is similar to previously proposed methods (46). Additionally, several targets were evaluated that were not fully supported by omics predictions or predicted as CsrA binding sites. While some of these target sequences were included to provide a potential pool of negative controls, we found that many of these targets had ambiguous omics data that could suggest a potential relationship to Csr under less stringent statistical methods. Although the evidence for these specific targets was not strong enough to be established as a target in the analysis done in our collaboration, these targets would have a statistically higher rate of interactions with CsrA than a truly random pool of targets. This distinction is important for the model based characterization of these targets performed in Chapter 4. The rationale for selecting a target mRNA appears in Table 3.1.

The mRNA sequences were designed with the shortest identifiable 5'UTR annotated by RegulonDB. In cases where genes had no designated transcriptional start sites or the gene was an internal gene in an operon, the 5'UTR was arbitrarily designated as 100 nucleotides preceding the translational start site.

A complete list of the mRNAs and the sequences tested can be found in Appendix Table A.1.

3.2.3 Selected targets display a wide range of reactions to CsrA

241 targets were tested in the 5' translational assay, of which 140 targets displayed measurable levels of GFP production that allowed us to evaluate their potential of being regulated by CsrA. The designed targets likely had a high failure rate due to the insolubility of the fusion protein produced. CsrA based repression or activation of targets was measured as the ratio of the background corrected average fluorescence output when CsrA was not induced (low levels of free CsrA) to the corrected fluorescence output when CsrA was induced (high levels of free CsrA). The background fluorescence was estimated as 3 fluorescence units (Section 3.4.2). Values greater than 1 indicate CsrA repression of the mRNA's translational activity. Table 3.1 reports the fluorescence values of the two replicates for when CsrA was induced and not induced, the ratio of the average fluorescence values, and the method that identified the gene as a potential target. A * symbol next to the average ratio indicates targets that had significantly variable replicate values compared to the ratio. Table 3.2 reports the targets that did not have sufficient fluorescence activities for further analysis.

Table 3.1: Fluorescence output of 5'UTR-CDS translation assay in relation to CsrA production

Gene name	Fluorescence with CsrA		Fluorescence without CsrA		Repression Ratio (No CsrA/CsrA)		Method of selection
glgC	11.00		362.50		44.94		Omics data
pckA	14.03	14.40	406.64	424.20	36.77		Omics data
glsA	83.18	28.73	933.21	1086.62	19.01		Omics data
aidB	32.27	40.71	563.78	575.95	16.93		Omics data
rseA	29.27	25.54	234.95	338.07	11.62		Motif prediction
maeB	13.90	14.30	109.01	146.14	11.22		Omics data
thiG	31.85	30.02	300.21	324.27	11.07		Omics data
yidQ	13.38	14.05	111.36	131.85	11.07		Motif prediction
yafQ	163.23	166.02	1755.13	1644.69	10.5		Motif prediction
mscS	12.39	12.09	79.66	111.60	10.02		Omics data
uspG	27.41	27.13	217.64	231.74	9.13		Omics data
deoD	23.76	21.96	145.77	214.89	8.93		Motif prediction
uidR	27.87	22.40	214.70	180.95	8.8		Motif prediction
acnA	9.05	8.99	44.30	56.89	7.91		Omics data
ydhQ	75.48	78.80	614.70	517.68	7.6		Omics data
frdB	126.54	117.65	887.13	895.14	7.46		Omics data
rnk	39.88	32.98	251.76	250.67	7.42		Omics data
hflK	82.82	90.03	546.82	690.03	7.38		Omics data
astD	42.78	33.10	249.20	248.99	7.04		Weak/no evidence
poxB	7.82		35.72		6.79		Omics data
nnr	50.25	109.46	390.51	638.23	6.65		Weak/no evidence
gshB	37.12	41.16	231.69	252.03	6.61		Motif prediction
fucO	678.09		4203.25		6.22		Omics data
moaB	45.94	42.39	231.33	254.66	5.83		Omics data
gstA	10.5	15.5	60.78	59.76	5.73		Omics data
ucpA	230.74	292.04	1434.58	1519.02	5.7		Omics data
cysD	9.28	16.81	50.98	66.12	5.53		Motif prediction
cmk	33.42	34.59	178.91	153.08	5.26		Weak/no evidence

Table 3.1 (continued)

Gene name	Fluorescence with CsrA		Fluorescence without CsrA		Repression Ratio (No CsrA/CsrA)		Method of selection
rnr	191.22	219.62	954.43	1002.76	4.82		Weak/no evidence
hemX	99.27	85.28	387.8	477.01	4.81		Omics data
ltaE	19.47	21.36	87.24	86.14	4.81		Omics data
glpR	47.44	47.89	214.89	219.32	4.79		Omics data
entF	11.59	12.07	43.24	45.43	4.68		Omics data
gadA	31.25	36.99	149.00	139.34	4.54		Omics data
ahr	127.33	130.72	606.55	484.99	4.31		Omics data
tnaA	6.51	7.22	21.65	17.2	4.25		Motif prediction
pgaA	10.56	12.46	34.51	42.17	4.15		Omics data
hfq	10.74	11.15	33.22	36.7	4.02		Omics data
entC	15.17	15.97	51.90	54.00	3.97		Omics data
clpB	27.68	25.17	99.07	90.09	3.91		Omics data
rpoS	31.00	31.78	107.54	115.92	3.83		Omics data
uxaB	48.61	51.52	177.32	183.65	3.77		Omics data
pgm	69.9	79.05	274.04	261.42	3.7		Omics data
tauD	87.62	93.59	304.57	337.22	3.63		Omics data
gadB	40.02	33.74	137.88	108.61	3.55		Omics data
proP	12.56	13.12	37.98	37.71	3.54		Omics data
dps	29.91	17.14	80.82	70.2	3.53		Omics data
relA	34.79	29.82	99.39	106.93	3.42		Motif prediction
icd	546.93	498.99	1756.53	1797.33	3.41		Motif prediction
ydeP	33.43	39.12	110.65	112.03	3.26		Omics data
evgA	9.00	9.44	24.37	21.64	3.22		Omics data
gstB	7.55	13.26	13.37	39.44	3.16	*	Weak/no evidence
sdhA	17.56	14.14	47.31	38.72	3.11		Omics data
groL	192.57	189.07	542.4	587.87	2.99		Omics data
ydjA	647.22	773.12	1985.81	2244.20	2.99		Omics data
rspB	7.32	6.90	13.60	16.51	2.93		Weak/no evidence
hemG	10.03	10.03	22.47	24.55	2.92		Omics data
yfgM	6.34	5.86	11.53	12.43	2.9		Motif prediction
nhaR	7.83	8.76	17.54	18.87	2.87		Omics data
csrA	52.08	48.55	142.48	120.23	2.71		Motif prediction

Table 3.1 (continued)

Gene name	Fluorescence with CsrA		Fluorescence without CsrA		Repression Ratio (No CsrA/CsrA)		Method of selection
sdiA	10.75	12.16	23.85	27.59	2.69		Known target
ybeL	89.81	82.98	162.32	280.18	2.62		Omics data
ybaL	9.16	9.10	17.76	18.97	2.51		Omics data
fdoH	7.13	7.28	12.87	13.71	2.45		Omics data
elaB	57.74	55.77	137.33	131.34	2.44		Omics data
crp	27.75	23.56	54.57	57.05	2.33		Weak/no evidence
cysJ	20.33	20.09	41.53	42.39	2.26		Omics data
cstA	25.99	29.71	53.12	63.22	2.22		Omics data
fabB	10.23	11.37	19.70	18.90	2.09		Motif prediction
dkgA	694.31	1319.67	1997.46	2105.04	2.04		Omics data
ppc	92.76	90.75	230.81	134.11	2.02	*	Motif prediction
pspA	33.41	32.25	62.98	63.50	2.02		Omics data
yhiI	9.78	11.40	19.29	16.64	1.97		Omics data
ahpC	1050.90	1884.68	2005.14	3731.33	1.96	*	Motif prediction
uxaA	67.36	61.08	117.37	125.22	1.93		Omics data
sucB	9.11	75.74	9.13	146.83	1.9		Omics data
ppk	20.01	20.82	37.50	33.98	1.88		Omics data
sucC	1040.28	1023.12	1734.97	1998.94	1.81		Omics data
yebE	627.76	619.18	1127.27	1100.09	1.79		Omics data
truC	7.37	7.26	10.67	9.78	1.67		Omics data
yaeP	15.49	14.54	22.31	23.21	1.64		Omics data
glcB	213.39	199.86	312.71	359.68	1.64		Omics data
dsrB	1287.09	1614.30	2242.49	2488.13	1.63		Omics data
cysK	196.12	185.57	273.41	337.99	1.61		Motif prediction
lsrF	563.76	639.45	914.99	1018.06	1.61		Omics data
ycaC	16.52	10.43	21.63	17.36	1.57		Omics data
iscS	179.04	238.30	363.66	279.00	1.55	*	Omics data
yecC	9.59	11.24	14.50	14.17	1.53		Omics data
katG	1107.23	1114.86	1690.52	1621.26	1.49		Weak/no evidence
bfr	48.35	57.03	72.84	75.50	1.43		Omics data
csiD	498.40	494.36	709.36	687.00	1.41		Omics data
rodZ	18.98	16.44	23.17	23.98	1.4		Weak/no evidence

Table 3.1 (continued)

Gene name	Fluorescence with CsrA		Fluorescence without CsrA		Repression Ratio (No CsrA/CsrA)		Method of selection
pepT	1499.42	1374.93	1894.61	2026.80	1.37		Weak/no evidence
groS	1174.81	1197.86	1551.13	1605.79	1.33		Omics data
iscR	184.19	193.26	252.90	244.29	1.32		Weak/no evidence
dnaK	611.82	736.98	869.26	848.69	1.27	*	Omics data
flu	183.48	80.91	209.58	122.13	1.26	*	Omics data
sucA	11.61	11.93	14.07	13.63	1.24		Omics data
aroG	229.99	216.58	285.83	257.30	1.22		Omics data
fabI	221.46	244.60	252.91	305.51	1.2	*	Motif prediction
carB	45.68	49.39	54.71	54.20	1.16		Omics data
yceD	45.53	37.81	46.22	47.64	1.14	*	Omics data
pntB	17.23	16.26	18.09	18.68	1.12	*	Omics data
proB	11.60	3.37	12.17	3.84	1.12	*	Weak/no evidence
fecA	42.75	38.93	44.06	45.31	1.10	*	Negative Control
hchA	520.13	482.02	529.41	560.37	1.09	*	Omics data
yjbD	12.94	14.79	14.82	14.79	1.09	*	Omics data
eno	28.58	26.97	29.60	29.80	1.08	*	Omics data
talA	1133.88	1247.33	1258.74	1256.47	1.06	*	Omics data
mreB	11.83	12.25	12.24	12.42	1.03	*	Omics data
uxaC	12.84	13.26	11.88	14.83	1.03	*	Omics data
pflB	96.68	94.55	85.93	109.98	1.03	*	Weak/no evidence
ydcS	540.39	525.45	540.32	530.08	1.00	*	Weak/no evidence
cdd	80.91	80.86	85.70	74.89	0.99	*	Omics data
yeaH	35.97	40.87	36.26	38.97	0.98	*	Weak/no evidence
sdhB	147.39	130.73	142.81	122.92	0.95	*	Omics data
wrbA	79.75	93.05	82.25	81.19	0.94	*	Omics data
pdxB	40.75	37.59	38.65	34.27	0.93	*	Omics data
metC	15.45	15.12	14.37	14.15	0.92	*	Weak/no evidence
lpxC	37.26	36.70	30.89	36.80	0.91	*	Weak/no evidence

Table 3.1 (continued)

Gene name	Fluorescence with CsrA		Fluorescence without CsrA		Repression Ratio (No CsrA/CsrA)		Method of selection
proS	101.06	87.90	86.41	82.71	0.89	*	Omics data
glnS	108.10	80.69	75.56	90.69	0.88	*	Weak/no evidence
katE	151.39	176.62	131.69	139.92	0.82	*	Weak/no evidence
ompR	13.90	13.66	11.82	11.64	0.81	*	Weak/no evidence
guaA	13.51	13.26	12.07	10.65	0.81	*	Omics data
kdsA	172.96	148.19	129.01	130.16	0.8	*	Omics data
aroD	32.52	34.45	25.31	28.35	0.78	*	Omics data
yqjD	2745.61	2558.64	1388.67	2678.87	0.77	*	Omics data
pta	13.28	12.96	10.50	10.98	0.76		Omics data
uxuB	41.52	37.47	31.77	29.17	0.75		Omics data
purM	179.72	180.67	125.67	140.53	0.73		Omics data
fhuA	11.82	11.07	9.51	8.89	0.73		Weak/no evidence
cspE	202.68	636.13	165.76	425.63	0.7	*	Weak/no evidence
iaaA	58.82	59.76	39.67	44.74	0.7		Weak/no evidence
glmS	14.18	13.00	10.38	10.21	0.69		Omics data
mtlD	506.11	506.49	341.94	313.97	0.65		Omics data
acs	2860.51	2987.32	1905.23	1823.10	0.64		Weak/no evidence
manX	24.66	28.91	15.67	18.32	0.59		Weak/no evidence
suhB	91.71	81.15	51.03	47.34	0.55		Weak/no evidence
ackA	9.66	47.35	6.72	7.60	0.16	*	Weak/no evidence

The mRNA product of each gene that produced fluorescence is displayed here. The fluorescence values are from the biological replicates of samples with induced or uninduced CsrA. Fluorescence values are in arbitrary fluorescence units and are relative to each other. The repression ratio is the background fluorescence corrected average of the two fluorescence conditions. A * symbol indicates targets that had significant variation between replicates as compared to the repression ratio. The final column describes the evidence that was used to select the target for the assay.

Table 3.2: mRNA targets with no observable fluorescence

Tested, not fluorescent							
adeP	*dacC*	*frdA*	*hdeB*	*purl*	*mlaA*	*yadM*	*yhcB*
adeQ	*dcrB*	*frvB*	*hflC*	*rbbA*	*nhaA*	*yaji*	*yhjG*
amyA	*ddlA*	*fumA*	*hipB*	*rib*	*nlpA*	*ybiT*	*yiaD*
arcA	*deaD*	*gabD*	*hlfC*	*rpoE*	*ntcA*	*ybjP*	*yihX*
argS	*dgcZ*	*galM*	*hyfR*	*rpsR*	*nuoC*	*ycaK*	*yiiS*
aroA	*dppA*	*gatC*	*isnH*	*sapA*	*nuoG*	*ycdT*	*yliE*
asd	*fbp*	*ghrA*	*ldcC*	*slp*	*oppA*	*ydcJ*	*yliF*
cbpA	*feoB*	*glgB*	*ldtA*	*sucB*	*pntA*	*ydgA*	*yqjE*
clpS	*fepA*	*glnH*	*lon*	*tamA*	*potD*	*yeaG*	*yqjG*
cyoA	*fes*	*gmk*	*lrhA*	*tgt*	*pqqL*	*yeaY*	*ytfQ*
cyoB	*flhC*	*gpt*	*maeA*	*thiM*	*ptsP*	*yebF*	
cysI	*flhD*	*hcaT*	*mdtA*	*topA*	*purE*	*yfhM*	
cysP	*fliY*	*hdeA*	*mdtE*	*tyrR*	*purK*	*yhbO*	

The maximum observable repression by CsrA was dependent upon the rate of cell division. Since GFP is such a stable molecule, the only way that GFP concentration is reduced is by cell division splitting the available pool of GFP by half for each division. For this reason, the maximum observed repression rate was 2^{-n}, with n being the number of cell divisions that took place. In this system, cell division was previously characterized as having roughly a doubling rate of 30 minutes without CsrA induction and between 30 and 60 minutes with CsrA induced (93). This allows for 3-6 cell divisions to occur within the 3 hours between CsrA induction and the fluorescence measurement (Section 3.4.2). This means that the maximum observable repression ratio was between 8 and 64. It should be noted that the doubling rate observed was highly variable between targets. Given this interpretation of repression, we observed that many mRNAs displayed nearly

total repression of translation, suggesting that CsrA is directly responsible for inhibiting the target.

Interestingly, not all targets that appeared to be inhibited by CsrA displayed total inhibition by CsrA; lower levels of repression were also observed. There are three possible scenarios to explain partial inhibition by CsrA: 1) indirect effects from CsrA expression are responsible for the observed inhibition, 2) CsrA interaction does not completely inhibit the target but instead creates a structural change in the transcript that alters translation rate, or 3) CsrA binding to the target mRNA is weak and incomplete, allowing there to be a mixture of bound and unbound transcript in the cell. The design of this assay was meant to mitigate the first scenario. The plasmid based system removes transcription from its natural regulation, but other post-transcriptional control mechanisms could be still effect the expression of the target. For example, CsrA controls expression of Hfq (36); any mRNAs that depend on Hfq would be indirectly affected by CsrA. The second scenario is likely true for some targets, but verification of this mechanism would require more characterization of the mRNA target. The third scenario is perhaps the most interesting possibility, as it has implications for the mechanism of CsrA based regulation. It is well documented that CsrA does not have equal affinity towards all mRNA targets, so the situation caused by weak CsrA affinity has to be true. This raises the question as to how much CsrA can be tuned naturally to affect different mRNAs at different cellular conditions. Differential affinities could create tunable mRNA activity. This mechanism for a global regulator is definitely a unique possibility that requires further exploration.

In addition to inhibition, several evaluated mRNAs also display the potential for activation by CsrA. As can be seen in Table 3.1, many genes display repression ratios significantly less than 1.0 (indicating that CsrA induction increased the observed fluorescence). While many of these targets were predicted to be activated by CsrA in our previously mentioned omics data, it is not possible to verify that these genes are truly activated by CsrA and not the result of indirect changes caused by CsrA. As mentioned previously, CsrA induction had the potential to alter the rate of cell division in some targets. If the cell division rate decreased while the translation rate of the gene remained constant, then the overall fluorescence observed would increase. In order to verify the relationship of CsrA and the mRNA target, there needs to be careful monitoring of cell division to ensure that the growth rate was not a major contributing factor to the increase in fluorescence intensity.

An important point to consider when evaluating the relationship between CsrA expression and fluorescence activity is that this test does not prove that CsrA is directly interacting with the mRNA. Although it is very likely that the most repressed targets are directly interacting with CsrA, it is possible that the observed repression could be the result of other unexpected indirect interactions. Due to the lingering concern that indirect actions are responsible for the observed behavior, it is necessary to prove that CsrA is directly interacting with the target mRNA. The remainder of this chapter presents our attempts to detect the direct interaction of CsrA with the target mRNA using modifications to the TriFC method presented in Chapter 2.

3.2.4 Modification of TriFC conditions to detect mRNA-CsrA interactions

In the TriFC method presented in Chapter 2, we developed a method that was capable of detecting the interaction of CsrA with the sRNA, CsrB, using molecular fusions to direct fluorescence complementation. In order to modify this system to be a diagnostic tool for mRNA-CsrA interactions, the RNA portion of the system had to be changed to produce the mRNA of the targeted genes.

The alteration of the RNA fusion for mRNA sequences has two major concerns: 1) the distance between a potential CsrA binding site and the MS2-binding site has to be close enough to promote refolding of the YFP fragments and 2) the mRNA has to be stable and properly folded to allow CsrA binding. These two concerns were less of an issue in the original system with CsrB because of the way CsrB binds CsrA. With regards to the first concern, CsrB has multiple sites that bind to CsrA, so the MS2-binding site is guaranteed to be near a CsrA binding site. Since the length of some mRNAs can be thousands of nucleotides, the location of the MS2BD needs to be considered carefully. The only location to attach the MS2-binding site is at the 5' UTR of the targeted mRNA. This is the most logical place, since CsrA binds in or near to the 5'UTR of regulated mRNAs.

The position of the MS2-binding site raises the concern that the transcriptional instability present in the MS2-binding region (section 2.2.4) would prevent production of the full transcript or create secondary structure unusual to the gene in question. To show that the targeted mRNA is being produced, the mRNA sequence of interest was designed to create a translation fusion with a red fluorescent protein (mStrawberry). The presence

of the fluorescent protein would indicate that the mRNA fusion was present and that it was capable of attracting a ribosome regardless of any potential structural changes to the mRNA. This is important to the stability of the mRNA, as translation is known to stabilize mRNA transcripts (94). The modifications to TriFC system and the plasmids responsible for producing the system are depicted in Figure 3.2

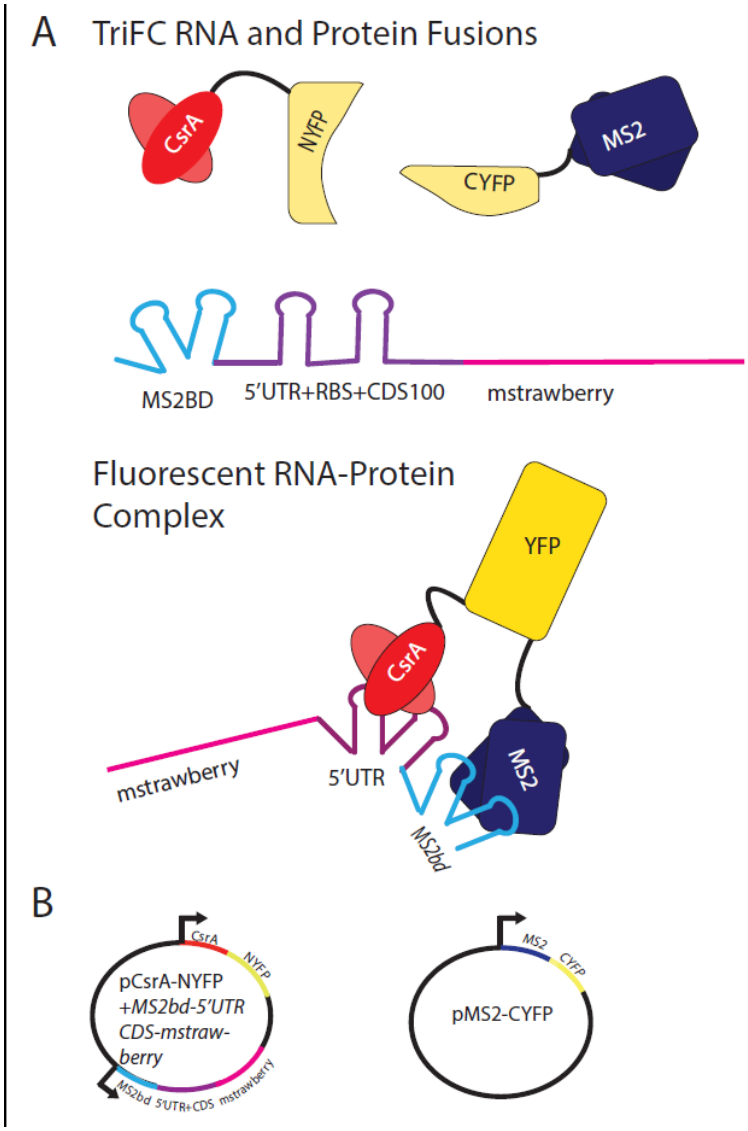


Figure 3.2: Modifications to CsrA TriFC for mRNA target identification

A) Schematic of CsrA TriFC – The two protein components used to create a fluorescent interaction are unmodified from the system discussed in Chapter 2. The RNA fusion is modified to create the MS2 binding domain at the 5' end of the mRNA target of interest. The mRNA produced a translational fusion with mStrawberry to produce a signal indicating the presence of the mRNA transcript.
B) The two plasmid system to produce the components of the TriFC system.

As a result, of the modifications to the TriFC system presented in Figure 3.2, cells produce two fluorescence signals, red and yellow. The yellow fluorescence is an indication of fluorescence complementation and is controlled by the interaction of CsrA and the mRNA target. The red fluorescence is the result of translation from the mRNA target.

3.2.5 TriFC provides affirmation of CsrA-mRNA interactions from translational assay

70 mRNA targets were selected to be tested by TriFC. Of these targets, 31 targets displayed increased yellow fluorescence over a negative control, indicating an mRNA-CsrA interaction. The negative control was the *fecA* mRNA which was shown not to be regulated by CsrA in 5'UTR assay (see Methods 3.4.6). Table 3.3 presents all of the mRNAs tested, the fluorescence emission corrected for cell concentration, the standard deviation, and the p-value (in relation to the negative control). Additionally, Table 3.3 includes an assessment of the data, indicating mRNAs that are likely directly interacting with CsrA. Any targets that displayed a higher average fluorescence than the negative control and also passed a Student's t-test with a value of $p<0.1$ were labelled as targets identified by TriFC. We used a slightly higher p-value because the fluorescence from TriFC had significant variation between replicates for some samples. If the fluorescence was significantly lower than the control, these targets were declared as having "No enhancement of TriFC." Targets that had failed the t-test were labelled as "evidence inconclusive."

It is interesting to note that many targets displayed fluorescence that was significantly lower than the fluorescence observed in the negative control. An explanation for this phenomenon is that CsrA may be attracted to the target, but CsrA binding is in a position that sterically hinders YFP complementation. Additionally, for some mRNA targets with highly active red fluorescence production, it is possible the mStrawberry could interfere with the observed yellow fluorescence by absorbing the emission spectra of YFP. In either case, it is important to understand that TriFC is only a positive test for CsrA binding to a target mRNA and that that a negative result does not indicate that CsrA is not interacting with a target. To illustrate this point, five mRNAs were tested that that had documented interactions with CsrA. These mRNAs were *glgC, pgaA, hfq, cstA*, and *nhaR*. While *glgC, nhaR*, and *pgaA* all displayed enhanced fluorescence, *cstA* and *hfq* failed to show enhanced fluorescence.

Table 3.3: Summary of TriFC expression for selected mRNA targets

mRNA	EM530/OD900	stdev	p-value	Outcome
fecA (*)	282	17	0.5	negative control
ahr	172	12	0.007	no enhancement of TriFC
acnA	491	145	0.072	Target
aidB	352	28	0.02	Target
amyA	349	31	0.027	Target
aroD	225	5	0.04	no enhancement of TriFC
bfr	164	34	0.008	no enhancement of TriFC
carB	126	22	0	no enhancement of TriFC
cdd	592	95	0.013	Target
clpB	1022	283	0.022	Target
csiD	235	37	0.087	no enhancement of TriFC
cstA	284	9	0.452	evidence inconclusive
cysK	157	10	0	no enhancement of TriFC
dps	1626	57	0	Target
dsrB	155	16	0	no enhancement of TriFC
elaB	195	22	0.003	no enhancement of TriFC
entF	387	218	0.245	evidence inconclusive
evgA	1011	219	0.014	Target
fbp	465	98	0.035	Target
fdoH	503	78	0.018	Target
fepA	255	61	0.293	no enhancement of TriFC
frdB	335	201	0.334	evidence inconclusive
gadA	229	53	0.112	no enhancement of TriFC
gadB	1033	314	0.027	Target
gadC	597	119	0.021	Target
glcB	169	32	0.006	no enhancement of TriFC
glgC	381	13	0.009	Target
glsA	881	112	0.005	Target

Table 3.3 (continued)

mRNA	EM530/OD900	stdev	p-value	Outcome
groL	390	119	0.127	evidence inconclusive
gstA	964	419	0.053	Target
guaA	131	21	0	no enhancement of TriFC
hemX	199	6	0.059	no enhancement of TriFC
hflK	226	39	6	no enhancement of TriFC
hfq	137	26	0.001	no enhancement of TriFC
kdsA	427	136	0.242	evidence inconclusive
ldtA	629	3	0	Target
lsrF	186	2	0.016	no enhancement of TriFC
ltaE	122	8	0	no enhancement of TriFC
moaB	278	82	0.472	no enhancement of TriFC
nhaR	548	36	0	Target
nuoC	276	66	0.451	no enhancement of TriFC
nuoG	173	22	0.001	no enhancement of TriFC
ompC	175	8	0.01	no enhancement of TriFC
osmE	267	100	0.412	no enhancement of TriFC
patA	488	8	0	Target
pck	216	78	0.141	no enhancement of TriFC
pgaA	703	271	0.057	Target
pgm	332	28	0.034	Target
phoU	160	7	0.008	no enhancement of TriFC
proP	858	393	0.063	Target
purM	274	114	0.461	no enhancement of TriFC
rib	340	17	0.03	Target
rpoS	453	59	0.014	Target
sdhA	494	9	0	Target
sucC	330	13	0.047	Target

Table 3.3 (continued)

mRNA	EM530/OD900	stdev	p-value	Outcome
tauD	176	22	0.002	no enhancement of TriFC
thiG	523	48	0.004	Target
uspG	485	146	0.067	Target
uxaC	241	84	0.25	no enhancement of TriFC
ybaL	336	34	0.045	Target
ybeL	145	22	0.002	no enhancement of TriFC
ybiT	344	39	0.046	Target
ydhQ	288	63	0.442	evidence inconclusive
ydjA	245	51	0.167	no enhancement of TriFC
yebE	620	62	0.004	Target
yecC	518	14	0	Target
yhjI	172	29	0.005	no enhancement of TriFC
yhjG	659	13	0	Target
yiiS	226	23	0.016	no enhancement of TriFC
yibD	514	68	0.011	Target
yqjE	302	5	0.172	evidence inconclusive
ytfQ	206	42	0.038	no enhancement of TriFC

mRNA targets were evaluated with TriFC by measuring the fluorescence emission at a wavelength of 530 nm (EM530). This value was normalized by the optical density at 900 nm (OD900) and an average value was taken for three measurements (see Methods 3.4.5). Values of fluorescence were compared to the *fecA* control (marked with a *). mRNAs that had statistically higher fluorescence (p-value) were labelled with an outcome of "target." mRNAs that failed the significance test were labelled as "evidence inconclusive." mRNAs that didn't appear to have enhanced fluorescence were labelled "no enhancement of TriFC."

3.2.6 Comparison of 5'UTR data TriFC complements translational assay data

In this assay, 32 targets were identified as having elevated TriFC readings as compared to a control. These targets are compared to the repression ratio from the translation assay presented in Table 3.1. The summary of this comparison appears in Table 3.5. While 8 targets are waiting for the data from the 5'UTR-translation assay (marked as "-"), only 2 genes did not display sensitivity to CsrA induction, *cdd* and *yjbD* (marked with a "*"). This means that the TriFC assay and the 5'UTR translational assay provide direct experimental evidence that CsrA directly binds to and regulates these 19 mRNA targets. This number excludes *glgC*, *pgaA*, and *nhaR*, which have been previously characterized as CsrA binding transcripts. While many of these targets have been suggested as CsrA targets based on previously published omics studies (21, 41), this work represents direct experimental evidence of CsrA both interacting with and having a regulatory effect on these targets.

Table 3.4: Comparison of TriFC targets and observed ratio change in fluorescence activity

mRNA	TriFC result	Repression ratio
cdd	target	0.99
yjbD	target	1.09
yecC	target	1.53
yebE	target	1.79
sucC	target	1.81
fdoH	target	2.45
ybaL	target	2.51
nhaR	target	2.87
sdhA	target	3.11
evgA	target	3.22
dps	target	3.53
proP	target	3.54
gadB	target	3.55
pgm	target	3.70
rpoS	target	3.83
clpB	target	3.91
pgaA	target	4.15
gstA	target	5.73
acnA	target	7.91
uspG	target	9.13
thiG	target	11.07
aidB	target	16.93
glsA	target	19.01
glgC	target	44.94
amyA	target	-
fbp	target	-
gadC	target	-
ldtA	target	-
patA	target	-
rib	target	-
ybiT	target	-
yhjG	target	-

This table displays the repression ration from Table 3.1 for the mRNAs that were identified as targets in the TriFC assay.

3.3 DISCUSSION

The work represents one of the most extensive characterizations of CsrA regulation with over 241 mRNAs evaluated for translational regulation by CsrA. Of these 241 targets, 94 targets displayed statistically significant changes in translation in relationship to CsrA. The modification to the TriFC tool developed in Chapter 2 created an *in vivo* system for the identification of mRNAs that interact with CsrA that provided evidence of 32 targets interacting with CsrA. With these two assays, a total of 19 previously unrecognized mRNAs were identified as CsrA targets.

Examination of the targets identified with these two assays has significant impact to the understanding of CsrA regulation. The implication that CsrA regulates *uspG* (the protein is a substrate of GroEL) and *clpB* (a protein folding chaperon) ties CsrA into the heat stress response. The regulation of *aidB* and *dps* shows that CsrA plays a role in responding to and preventing damage to DNA. The regulation of *evgA* (a transcriptional regulator of acid response genes) ties CsrA into regulation of acid stress response, which explains previous findings that CsrA had a relationship to the acid response in *H. pylori* (95). The regulation *sdhA* and *sucC* (genes that produce succinate) show how CsrA is involved in the citric acid cycle and even suggests that modifications to the Csr system could be a productive strategy to improve the fermentation of succinate (96).

Most notably, the sigma factor r*poS* appears to be inhibited by CsrA. As a sigma factor, RpoS is responsible for transcriptional control of numerous genes with specific control over stress response genes and stationary phase genes. As these two functions are responsible for major cellular changes, the finding that CsrA controls RpoS concentration

provides an explanation as to how CsrA can be involved with so many genes. Additionally, it is known that RpoS is partially responsible for transcription of *csrA* (97), the implication that CsrA inhibits its activators suggests that the cell can carefully titrate CsrA expression to meet certain conditions.

While the total number of targets verified by TriFC is much smaller than we had hoped, it is important to clarify that the TriFC test only provides affirmation of targets interacting with CsrA. Since the test is expected to fail to identify targets due to steric factors associated with the position of CsrA binding, a negative result for TriFC binding does not indicate that there is no interaction between CsrA and the target. Given that the 94 targets that responded to CsrA expression display a wide range of relationships to CsrA activity, including what can be interpreted as total inhibition of translation, it is likely that many of these genes are actually regulated by CsrA. Furthermore, the number of targets analyzed in this translational assay represents the largest biophysical-mechanistic characterization of the Csr system. In the following chapter, we present a model that interprets the findings of this mechanistic characterization of CsrA and has implications for CsrA function.

3.4 METHODS

3.4.1 Plasmid creation for translational Assay

The translation assay presented here is a modified system of translation assay presented in previously published work (93). Selected target sequences were created by genomic PCR amplification off of MG1655 *E. coli* or produced through Genscript DNA synthesis services. Selected targets were inserted into plasmid pHL1756 between the SalI

and SphI cut sites to create the test plasmid. Two methods were used to create the pHL1756 variants, either standard restriction site cloning or Gibson assembly. The plasmids, pHL1756 and pHL600, and the strain HL4142 are all available through Addgene ® services. A complete list of primers used to amplify genomic DNA is found in Appendix Table A.2.

3.4.2 Growth conditions and fluorescence measurement for translational assay

To perform the experiment, cells from the strain HL4142 were transformed with the plasmid pHL600 followed by a second transformation of the gene specific variant of pHL1756 (both using a $CaCl_2$ chemical transformation technique). Colonies of cells containing both plasmids were grown overnight to be used to inoculate two separate 250mL flasks containing 10 mL of fresh LB media. Overnight culture was added to the fresh media at a ratio of 1:50. These cells were grown at 37°C with agitation to an approximate OD600 of 0.3, which typically took between 2-3 hours. At this time, one of the two inoculated flasks was induced to express CsrA by adding IPTG to a final concentration of 0.1 mM. This set up compares the conditions of CsrA induced and uninduced. Each mRNA target was tested with biological duplicates.

Fluorescence was measured 3 hours after inducing CsrA by flow cytometry on a FACSCalibur flow cytometer (BD Biosciences). This amount of time was chosen in order to balance the need to limit the accumulative effects of indirect regulation caused by CsrA induction with the need to allow enough cell divisions to dilute GFP concentrations (GFP is very stable). The fluorescence ratio has been corrected for the background fluorescence by subtracting 3 fluorescence units from the average value. The

background fluorescence is an approximated value because there was considerable variation around this value between day to day measurements. The lowest fluorescence value observed varied between 3 and 8 units (the lowest reading was actually 2.5 for some inactive targets). The distinction between 3 and 8 units could produce significant changes in the observed ratio of repression for targets with weaker fluorescence. However, the uncertainty around the background fluorescence only affects a small portion of targets.

3.4.3 Identification of targets effected by CsrA induction

The ratio of the mean fluorescence of the population expressing CsrA relative to cells not expressing CsrA was calculated to determine if CsrA expression altered translational activity of the mRNA. Constructs were tested in biological duplicate to verify that the observed trends were reproducible and statistically significant. If the Student's t-test comparing an mRNA's fluorescence with CsrA induced and non-induced produced a p-value less than 0.1, that target was labelled as affected by CsrA. Targets that failed the t-test were marked by the * symbol in Table 3.1.

3.4.4 Construction of TriFC plasmids

The TriFC plasmids used in the system were derived from the plasmids used in our previously published work with CsrA-CsrB TriFC (98). The MS2-CYFP expressing construct is unmodified from the original plasmid presented in Chapter 2. The test plasmid pCsrA-NYFP+MS2bd-5'UTRCDS-mstrawberry was constructed from that study's plasmid pCsrA-NYFP+2MS2bd (presented as a control in Chapter 2). This plasmid was modified by the insertion of the sequence speI-RBS-sphI-mstrawberry-aatII

purchased from Genscript (Appendix A.3). The sequence was inserted into the plasmid between the SpeI and AatII cut sites following the 2MS2bd sequence. This sequence introduced a ribosome binding site and a start codon that would produce mStrawberry. Additionally, a SphI cut site was created preceding the *mStrawberry* sequence so a 100 nucleotide length of coding sequence inserted at the SphI site would be in frame with the sequence of *mStrawberry*.

The mRNA sequences of selected genes tested via TriFC were identical to the sequence tested in the translation assay. Sequences were created by through Genscript ® or from a genomic PCR. All sequences were inserted into the plasmid pCsrA-NYFP+MS2bd-5'UTRCDS-mstrawberry at the restriction sites of SpeI and SphI. All sequences were verified by the ICMB Core Facility sequencing facility available at the University of Texas at Austin. For a complete list of targets tested with the sequence used, please see Appendix A.1.

3.4.5 Test conditions for TriFC

The TriFC plasmids and the MS2-CYFP containing plasmid were transformed into a strain with no native expression of CsrB, *ΔcsrB* trMG1655 (courtesy of Tony Romeo (25)). The strain was chosen to prevent CsrB from interfering with the TriFC interaction. This is identical to the strain used in Chapter 2.

To create the two plasmid system needed for TriFC the strain *ΔcsrB* trMG1655 is transformed with two sequential transformations. The plasmid pMS2-CYFP is transformed into the strain first using $CaCl_2$ chemical transformation methods. Cells containing pMS2-CYFP are made competent and transformed with the TriFC plasmid

(also using $CaCl_2$ chemical competent transformation). The transformed cells are grown on plates carrying the appropriate antibiotic conditions (carbenicillin and kanamycin).

For each mRNA sequence tested, three colonies were picked and inoculated into 5mL of LB in the presence of carbenicillin and kanamycin at concentrations of 100 µg/mL and 50 µg/mL, respectively. The cells were grown for 18 hours after inoculation. It was not necessary to induce the expression of the TriFC constructs because the *ΔcsrB* strain constitutively expressed the constructs (discussed in Chapter 2). At this point, the cells are harvested and measured for fluorescence.

The cells are harvested by pelleting 1 mL of saturated liquid culture in 1.5 mL conical centrifuge tube. The LB supernatant is decanted and the cells are re-suspended with 1 mL of phosphate buffered saline solution (PBS) and pelleted a second time followed by removal of the PBS supernatant. The cells are re-suspended again in 1 mL of PBS. Fluorescence is measured by plate reader using the Cytation3 Imaging Reader. 300 µL of the washed cell culture is loaded into a 96 well plate that is black with clear bottom (Corning Costar Assay Plate). Fluorescence is excited with a 470 nm laser wavelength, and the fluorescence emission is measured at 530 nm using spectral scanning. This measurement is normalized by reading the optical density at 900 nm (OD900). The wavelength of 900 nm was chosen to avoid potential interference by the mStrawberry protein that could occur with a traditional 600 nm optical density reading. The fluorescence was reported as fluorescence units per OD900 units, with OD900 being corrected using a PBS blank.

3.4.6 Establishing a fluorescent control for system

To establish a fluorescence control, mRNA targets were chosen from the currently unpublished omics data referenced in "Target selection" that displayed little relationship to CsrA. These mRNAs were tested in the 5'UTR translation assay to assess their sensitivity to CsrA, and it was determined that the gene *fecA* was not sensitive to CsrA. Since it was expected that *fecA* would not interact with CsrA, the fluorescence from *fecA* TriFC should represent the background recombination of YFP. To compare different targets to *fecA*, the fluorescence emission spectra of several mRNA targets were measured as described in the Section 3.4.5 of the methods. The spectral intensity of select mRNAs appears in Figure 3.3.

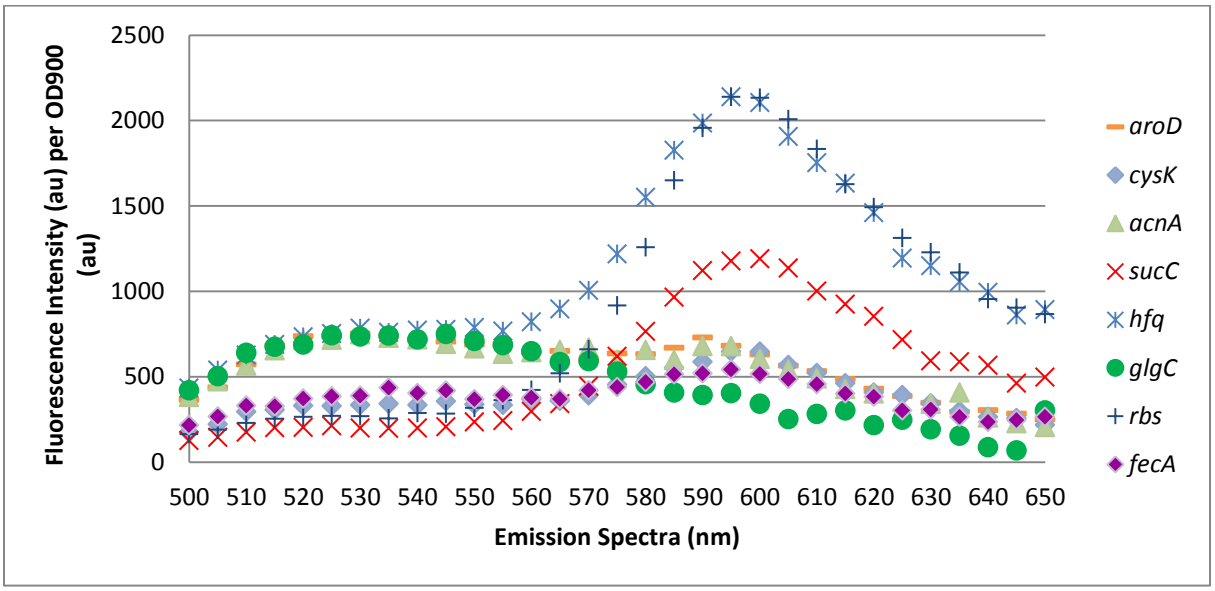


Figure 3.3: Spectral fluorescence emission of selected targets in the TriFC system

The curves are of the fluorescence intensity measured at specific wavelengths for different genes being tested in the TriFC system. The two peaks in the spectra relate to the two fluorescent proteins present in the system, with the peak at 530 nm coming from refolded YFP fragments and the peak at 600 coming from mStrawberry.

In Figure 3.3, the two fluorescence proteins are clearly visible by the presence of an emission peak at 530 nm, corresponding to recombined YFP, and at 600 nm, corresponding to mStrawberry. The targets *aroD, acnA, glgC* and *hfq* display significantly higher fluorescence at 530 nm than the targets *fecA, sucC, cysK*, and an artificial RBS control (mentioned in 3.4.4). The interpretation of this result is that these targets are enhancing yellow fluorescence as a result of CsrA-target interactions. It should be noted that *hfq* and *glgC* are documented targets of CsrA, which reinforces the interpretation that CsrA is directing fluorescence complementation. Additionally, it can be seen that the intensity of the fluorescence from mStrawberry varies with respect to the targets being tested. The mStrawberry fluorescence is dependent upon the translational activity of the target gene, which could be affected by CsrA regulation. However, there appears to be no correlation between yellow fluorescence and red fluorescence, suggesting that the TriFC measurement is independent of red fluorescence.

To establish the statistical significance of the fluorescence measurements taken, all targets were tested in biological triplicates. Using the *fecA* target as a control, the Student's t-test was performed for all targets with an average yellow fluorescence value greater than that of *fecA*. For the purposes of identifying targets that show enhanced TriFC, a p value of 0.1 was used as a cut off. The emission summary of the targets displayed in Figure 3.3 appears in Table 3.4.

Table 3.5: Summary of fluorescence emission from triplicates of targets presented in Figure 3.3

Gene Target	Average Fluorescence 530 nm normalized to OD900 (arbitrary units)	Standard Deviation of 530 nm fluorescence	p value using fecA as a control	Average Fluorescence 600 nm normalized to OD900 (arbitrary units)	Standard Deviation of 600 nm fluorescence
aroD	653	52	0.002	610	32
cysK	309	12		556	26
acnA	722	214	0.072	792	233
sucC	401	203		521	101
hfq	405	317		1190	761
glgC	682	106	0.01	300	49
rbs	235	10		1893	87
fecA	415	45		529	39

As can be seen in the comparison shown in Table 3.5, the targets *aroD*, *acnA*, and *glgC* have a significantly larger amount of yellow fluorescence as compared to the *fecA* control. The other replicates of the known CsrA target *hfq* were not as fluorescent as the replicate depicted in Figure 3.3. This means that the target *hfq* can not be declared a target using TriFC data because it fails to past the Student's t-test for significance. Expanded testing of *sucC* suggested that fluorescence presented in Figure 3.3 is indeed an outlier and that *sucC* is probably a target (presented in Results 3.2.4). The issue with *hfq* and *sucC* is representative of the TriFC system. Some mRNAs will display a wide range of fluorescence capabilities between biological replicates, making it difficult to determine if the gene is interacting with CsrA by standard statistical methods. It is likely that mRNAs that display enhanced yellow fluorescence inconsistently are likely CsrA targets that have a complex CsrA-target relationship.

For the purposes of comparing measurements taken on different days, all measurements are normalized to a *fecA* control.

Chapter 4

A first principles model of CsrA binding and activity with experimental validation

4.1 Introduction

As computational methods advance in sophistication and efficiency, researchers turn to *in silico* methods to create models capable of informing and predicting biological behavior. For regulatory RNAs, most modelling efforts are focused on sRNA-mRNA interactions and have largely ignored other regulatory RNAs. These sRNA-mRNA models benefit from the understanding of base-pair interactions and focus largely on the prediction of mRNA targets (67, 99-101). These models make no effort to predict the extent of the regulatory interaction, but they are valuable for directing more in depth characterizations of sRNA-target interactions. Currently, few models exist for other regulatory RNA interactions due to the lack of a mechanistic understanding of the specific interactions.

Compared to the well understood base-pair interactions of sRNA-mRNA regulation, there are no generalizable strategies for the modelling of RNA-protein interactions. Each RNA-protein interaction must be characterized individually to develop the knowledge to predict the system's behavior. Characterizations of RNA-protein interactions utilize the techniques of SELEX (54), SEQRS (102), and CLIP (51) to identify the protein's RNA recognition motif. However, these methods have been criticized as being biased towards identifying the highest affinity targets and potentially missing biologically relevant lower affinity targets (52). The quality of the interaction

motif characterization is of great importance because the motif can be used for target predictions similar to methods using sRNA motifs to predict mRNA targets.

With regard to the protein CsrA, SELEX and CLIP experiments have verified that CsrA recognizes GGA motifs, AAGGA specifically, with high affinity (29, 41) . The GGA motif has been used to make simplistic predictions of CsrA's mRNA targets that are based on either the presence of the GGA motif near the RBS or the prevalence of the motif in the transcript (21, 46). In the former case, hundreds of mRNA targets are predicted to interact with CsrA, which is likely too broad of a prediction. Conversely, the method that identifies targets based on the prevalence of the motif is likely too selective, identifying only a specific set of targets. Furthermore, these predictions formulated from sequence similarities to the recognition motif are solely qualitative and lack quantitative information that could explain CsrA behavior on these mRNAs.

In this work, we developed a biophysical model that predicts CsrA binding interactions and provides an assessment of the affinity of the interaction. The predictions of this model are interpreted through a published biophysical interpretation of translation initiation (the RBS Calculator) (103) in order to predict the regulatory significance of the CsrA-RNA interaction. This model's predictions are validated with experimental evidence of regulation based on the translational assay presented in Chapter 3. Furthermore, the quantitative nature of the model was investigated with experiments that demonstrate *in vivo* titration of specific mRNA transcripts. The combination of the binding site prediction and the effect it plays on translation presents a novel model to predict an mRNA interaction's potential with CsrA. Beyond target predictions, this

model provides a qualitative prediction of the affinity of mRNA-CsrA interactions that creates a rational basis to explain how CsrA fine-tunes expression of multiple targets.

4.2 RESULTS

4.2.1 CsrA binding model

The free energy model presented in this chapter was created from the first principles describing CsrA-RNA interactions and predicts the affinity of CsrA with any given sequence of RNA. This model makes four basic calculations to predict the most likely site for CsrA to bind to and predicts how binding at this site affects translation initiation (Figure 4.1). These steps ultimately determine: 1) all the minimum free energy structures for the RNA bound to CsrA and 2) the translation initiation rates of each of these structures. In this model, the RNA sequence analyzed is the 5'UTR and the first 100 nucleotides of the coding sequence of an mRNA. For comparison purposes, the 5'UTRs evaluated in the model are identical to those used experimentally in Chapter 3.

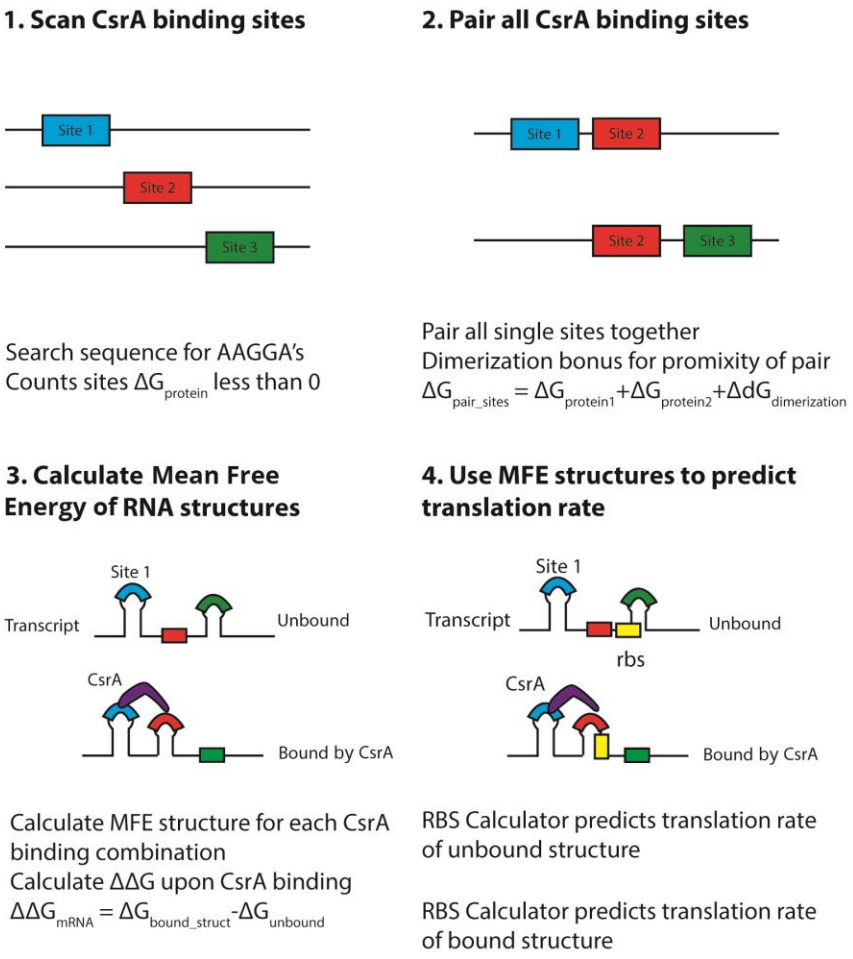


Figure 4.1: Calculations to determine CsrA affinity and effect on translation

These four panels represent the four steps that the model uses to predict the position, affinity, and regulatory effect that CsrA will have on an mRNA sequence.

Step 1 – The model scans the sequence for 5 nucleotide fragments that have a similarity to the AAGGA motif and assigns an initial free energy score according to the position weight matrix.

Step 2 – The model calculates the free energies of all pairs of sites identified in Step 1. An additional free energy term is added based on the distance between the sites.

Step 3 – The model calculates the minimum free energy of the RNA structure that will permit CsrA to bind the sites identified in Step 2.

Step 4 – The model uses the predicted RNA structure of the transcript bound by CsrA and free of CsrA to modify the RBS Calculator to predict the overall effect that CsrA binding has on translation.

The first step of the model is to determine primary CsrA binding sites. In bacteria, CsrA binding sites have been characterized as containing an ANGGA motif (29, 31, 32). Specifically, SELEX experiments defined the motif for *E. coli* CsrA as AAGGA and identified the changes in K_d that occurred when each nucleotide in the AAGGA motif is mutated. (29). From this experiment, a position weight matrix (PWM) was made to describe the ΔG associated with each nucleotide in the motif (Table 4.1).

Table 4.1: Position weight matrix of CsrA binding site

A	U	G	C
-2.63	0	0	0
-2.20	0	0	0
0	0	-3.14	0
0	0	-3.14	0
-1.65	0	0	0

Each nucleotide in a sequence that matches the sequence AAGGA is assigned an individual free energy contribution according to its position in the motif.

This PWM creates estimations for CsrA affinity ($\Delta G_{protein\_binding}$) for every set of 5 nucleotides in a sequence. This method of identifying CsrA binding sites also broadens the search beyond the traditional standard of requiring a GGA sequence within the motif (21, 45). It is important to note that the experiment defining the PWM only performed one mutation of the AAGGA motif, so it is only possible to calculate the free energy contributions of AAGGA and no other nucleotides that may be able to contribute. In addition, other efforts using molecular modelling to determine the energetic contributions of individual nucleotides have determined slightly different energetic values (31).

Using the PWM, the full sequence of a given UTR was scanned for five nucleotide sequences with $\Delta G_{protein\_binding}$ less than 0, and all sites that met this criterion were identified and sorted by lowest to highest $\Delta G_{protein\_binding}$ (Figure 4.1-1).

Since CsrA is known to bind multiple sites on a transcript (19, 37), the model for likely CsrA binding sites was extended to include free energy estimations for CsrA binding two sites simultaneously. The minimum free energy of all unique pairs of single CsrA sites was calculated as the sum of their individual ΔG's of CsrA binding (Figure 4.1-2).

$$\Delta G_{pair\_sites} = \Delta G_{protein\_binding\_1} + \Delta G_{protein\_binding\_2}$$

In addition to the consideration that CsrA can potentially bind to two RNA sites, previous research observed that pairs of sites often bind to the same CsrA when the distance between the sites is between 10 and 50 nucleotides (32). This observation suggests that RNA sites bind to CsrA cooperatively. To account for this cooperativity, the term $\Delta G_{dimerization}$ was added to the model to provide an additional ΔG bonus for CsrA binding sites between 10-50 nucleotides apart. The $\Delta G_{dimerization}$ is calculated with the following equation:

$$\Delta G_{dimerization} = 0.001d^2 + 0.05d - 5.0$$

where d is the number of nucleotides between the two predicted sites

when $d \leq 11$, Steric penalty of $\Delta G = 20$

Representing dimerization in this way provides a diminishing dimerization bonus for sites as they become more separated (with no bonus given after 50 nucleotides).

The model calculates the minimum free energy of a possible CsrA interaction as sum of the single CsrA binding sites identified by the PWM and the dimerization term. The free energy of CsrA binding to a pair of sites is depicted below:

$$\Delta G_{pair\_sites} = \Delta G_{protein\_binding\_1} + \Delta G_{protein\_binding\_2} + \Delta G_{dimerization}$$

The presented model for CsrA binding makes two major assumptions for how CsrA interacts with RNA. The first assumption is that CsrA binds to two sites and each site has an accumulative effect, and the second assumption is that the distance between these sites can positively influence the affinity of the interaction. As part of the evaluation of this model, we decided to evaluate the assumptions to determine if they were meaningful to the physical interaction. This was accomplished by altering the free energy model to neglect the contributions of either the dimerization term or the second site. Effectively, this created three different models that predicted different CsrA binding sites. The first model assumes that CsrA only binds to one site on the mRNA and recognition depends exclusively on the PWM (Single site model). The second model allows CsrA to bind to two sites, but CsrA receives no bonus from the relative position of the two sites (Two site model). Finally, the third model is the previously described model (Dimer model). These three models created different ΔG predictions for predicted sites which changed the prediction of the most likely binding site. These three models are evaluated in the following section (Section 4.2.2).

Once all possible combinations of CsrA binding sites had been determined using the three different model assumptions, a minimum free energy structure was calculated for all CsrA binding sites predicted using ViennaRNA (104, 105) (Figure 4.1-3). The

structure of CsrA bound mRNA was defined as the most stable RNA structure that leaves the predicted binding sites free of secondary structure. The structure of the 5' UTR unbound by CsrA was calculated and used to predict the energy change that results from the change in RNA structure caused by CsrA binding. The energy change of the RNA structure bound to CsrA is presented below:

$$\Delta\Delta G_{mRNA} = \Delta G_{bound\_structure} - \Delta G_{unbound}.$$

This energy change based on the RNA structure was the final term used to calculate the overall ΔG of CsrA binding. Depending on the model assumptions being used, the free energy predictions for CsrA binding sites were based on the similarity of a site (or sites) to the PWM, the distance between that site and a second site, and the structural change that occurs to allow CsrA to bind at those positions. The most likely predicted structure (lowest free energy) was input into the RBS Calculator (106) to determine the translation initiation rate of CsrA bound mRNA (Figure 4.1-4).

For unbound mRNA transcripts, the RBS Calculator predicted a ΔG_{total} of translation initiation using the following equation:

$$\Delta G_{total} = (\Delta G_{mRNA:rRNA} + \Delta G_{start} + \Delta G_{spacing} - \Delta G_{standby}) - \Delta G_{mRNA\_structure}$$

Where $\Delta G_{mRNA:rRNA}$, ΔG_{start}, $\Delta G_{spacing}$, and $\Delta G_{standby}$, are ΔG terms that described mRNA binding to the ribosome, and the term $\Delta G_{mRNA\_structure}$ is the energy required to alter the RNA structure to allow ribosome binding (106). The RBS Calculator calculated the translation initiation rate (TIR) using a partition function relating ΔG_{total} to TIR:

$$TIR = e^{-(\beta \Delta Gtotal)}$$

Where the Boltzmann constant, β, was previous determined to be 0.45 ± 0.05 mol/kcal in previous studies (106, 107).

The structural consideration for CsrA binding was introduced into the equation through modification to the $\Delta G_{mRNA\_structure}$ term to create the new term $\Delta G_{bound\_structure}$. This term accounts for the energy change required to alter the structure of the RNA bound by CsrA to allow ribosome binding and is relative to the original $\Delta G_{mRNA\_structure}$. Functionally, the modification fits into the ΔG_{total} as follows:

$$\Delta G_{total} = (\Delta G_{mRNA:rRNA} + \Delta G_{start} + \Delta G_{spacing} - \Delta G_{standby}) - \Delta G_{bound\_structure}$$

Since the $\Delta G_{bound\_structure}$ is relative to the $\Delta G_{mRNA\_structure}$ term, it can be positive, negative, or neutral with regards to the ribosome binding. If CsrA binding to a transcript causes the RNA to adopt a structure that blocks the ribosome, the term is negative (reduces the predicted translation rate). A neutral term means that CsrA binding the RNA causes no change to the structure. For a positive $\Delta G_{bound\_structure}$, CsrA binding opens up the RBS site and would activate translation.

In addition to $\Delta G_{bound\_structure}$ factoring in the structural changes in the RNA due to CsrA binding, an additional reduction to the translation rate was added to the model if CsrA was predicted to bind on the ribosome binding site (RBS). This condition is described as TIR_{on} and is calculated below.

$$TIR_{on} = TIR/e^{(\beta \Delta Gprotein\_site\_1)}$$

$\Delta G_{protein\_site\_1}$ is the ΔG derived from the PWM for the 5 nucleotide binding site

This calculation of TIR_{on} effectively alters the total free energy calculation to include a term for the energy to remove CsrA from a single position in the following manner:

$$\Delta G_{total} = (\Delta G_{mRNA:rRNA} + \Delta G_{start} + \Delta G_{spacing} - \Delta G_{standby}) - \Delta G_{bound\_structure} - \Delta G_{protein\_site\_1}$$

CsrA binding sites that occur within the RBS sequence result in a significant reduction of translation. In this model, it is assumed that only one CsrA position will directly interfere with translation, even if both sites are positioned to do so.

With the consideration of the RNA structure and the position of CsrA binding, the RBS Calculator provided an estimate for the translation rate of the mRNA when CsrA is bound and when it is not bound.

It is important to understand that this model predicting the activity of CsrA on mRNA has two parts. The first part predicted an ensemble of potential CsrA binding sites, but the second part predicted how CsrA binding would affect the translation rate at only the most likely site. This is done for practical reasons. Given that the model predicts all potential binding sites, there will usually be one binding site combination that will inhibit translation. However, most of the predicted sites are not relevant under biological conditions, so the model only considers the most likely binding site. The effectiveness of this model and the assumptions made in predicting the binding sites is discussed in the following sections.

4.2.2 Verification of the model using translation assay data

This model provided two predictions: 1) the position on the mRNA that is most likely to bind CsrA with a relative estimate of affinity and 2) the effect that CsrA binding will have on translation. While the verification of the binding site of CsrA would require extensive *in vitro* experimentation, the 5' translation assay presented in Chapter 3 would be the ideal method to evaluate the prediction of CsrA's effect on translation. To evaluate our model with the data from the translation assay, it was necessary to make two major assumptions: 1) Any mRNA activity that was repressed at a minimum ratio of 1.5 (fluorescence without CsrA/fluorescence with CsrA) is directly interacting with CsrA and 2) the genes tested are representative of the whole population of interacting and non-interacting genes. With these considerations, all 5'UTR sequences that had measurable fluorescence values were evaluated using the three different CsrA binding assumptions discussed in Section 4.2.1.

By evaluating the three models of CsrA binding site predictions, we were able to determine which factors are the most important to describing CsrA behavior. The first model assessed assumes that CsrA binds to only one position (named the Single site model). The second model allows for CsrA binding at two sites and neglects the value of dimerization (named the Two sites model), and the third model includes the estimation of dimerization (Dimer model). Table 4.2 presents the prediction of the ΔG of CsrA binding for each model, the prediction of translation activity with CsrA for each model, the observed ratio of repression by CsrA (fluorescence without CsrA/fluorescence with

CsrA), and the predicted ratio of fluorescence taken from the RBS Calculator predictions (only for the Dimer model of CsrA binding).

Table 4.2: Model predictions for CsrA activity on specific mRNA

mRNA Target	Average Free Fluo. (au)	Measured Repression Ratio (No CsrA/CsrA)	Predicted Repression Ratio (No CsrA/CsrA)	Dimer Model ΔG CsrAbind	Dimer Pred	Two Sites ΔG CsrA bind	Two Sites Pred	Single Site ΔG CsrAbind	Single Site Pred
glgC	102.7	44.94	11884.95	-24.77	1	-20.18	1	-12.76	1
pckA	415.4	36.77	46.37	-22.65	1	-19.13	1	-10.40	1
glsA	1009.9	19.01	1.00	-21.81	2	-18.88	2	-10.56	2
aidB	569.9	16.93	94.80	-19.51	1	-15.93	2	-9.10	2
rseA	286.5	11.62	135.98	-24.42	1	-20.18	1	-10.56	2
maeB	127.6	11.22	1298930.45	-23.06	1	-18.47	1	-9.94	1
thiG	312.2	11.07	84.51	-15.78	1	-13.17	1	-6.99	2
yidQ	121.6	11.07	1.00	-25.34	2	-23.87	2	-12.76	2
yafQ	1699.9	10.50	304.18	-24.95	1	-21.27	1	-11.11	1
mscS	95.6	10.02	2289764.33	-23.58	1	-19.23	1	-10.32	1
uspG	224.7	9.13	199.52	-20.99	1	-17.74	1	-9.97	1
deoD	180.3	8.93	499359.43	-24.69	1	-20.05	1	-11.14	1
uidR	197.8	8.80	1.00	-21.12	2	-16.68	2	-8.71	2
acnA	50.6	7.91	0.23	-21.00	2	-17.15	1	-9.59	2
ydhQ	566.2	7.60	42244.48	-23.96	1	-20.28	1	-11.80	1
frdB	891.1	7.46	16456.77	-26.82	1	-22.28	1	-12.76	1
rnk	251.2	7.42	2112.61	-21.87	1	-18.39	1	-9.62	2
hflK	618.4	7.38	31059.50	-24.27	1	-20.96	1	-10.56	1
astD	249.1	7.04	1.00	-20.79	2	-17.80	1	-9.62	2
poxB	35.715	6.79	47.12	-20.20	1	-17.98	2	-10.56	1
nnr	514.4	6.65	477.05	-21.39	1	-17.37	1	-9.62	2
gshB	241.9	6.61	6615.68	-24.83	1	-20.18	1	-10.56	1
fucO	4203.3	6.22	2398.88	-26.72	1	-22.38	1	-12.76	1
moaB	243	5.83	9.17	-23.98	1	-20.18	1	-12.76	1
gstA	60.3	5.73	25.84	-22.07	1	-20.18	2	-12.76	2
ucpA	1476.8	5.70	4.41	-23.64	1	-22.11	1	-11.95	1
cysD	58.6	5.53	5782.56	-24.10	1	-19.86	1	-10.24	1
cmk	166	5.26	105776.28	-23.62	1	-21.37	1	-10.81	1
rnr	978.6	4.82	51.33	-20.12	1	-15.94	1	-7.97	1
hemX	432.4	4.81	621513.45	-22.96	1	-19.24	1	-9.62	1
ltaE	86.7	4.81	17225.49	-22.10	1	-18.94	1	-11.30	1
glpR	217.1	4.79	309.86	-20.89	1	-17.91	2	-8.77	1
entF	44.3	4.68	618517.85	-20.50	1	-15.76	1	-8.16	1

Table 4.2 (continued)

mRNA Target	Average Free Fluo. (au)	Measured Repression Ratio (No CsrA/CsrA)	Predicted Repression Ratio (No CsrA/CsrA)	Dimer Model ΔG CsrAbind	Dimer Pred	Two Sites ΔG CsrA bind	Two Sites Pred	Single Site ΔG CsrAbind	Single Site Pred
gadA	144.2	4.54	308278.76	-21.39	1	-17.67	1	-10.16	1
ahr	545.8	4.31	48478.28	-25.65	1	-22.38	1	-12.76	1
tnaA	19.4	4.25	7922232.31	-25.55	1	-20.90	1	-10.34	1
pgaA	38.3	4.15	3.24	-27.06	1	-22.36	1	-11.45	1
hfq	35	4.02	10493.75	-26.81	1	-22.57	1	-12.76	1
entC	53	3.97	0.27	-19.48	2	-16.88	2	-9.62	2
clpB	94.6	3.91	7.38	-17.77	1	-13.37	2	-6.89	2
rpoS	111.7	3.83	1.00	-20.88	2	-16.93	2	-8.61	2
uxaB	180.5	3.77	1.00	-20.27	2	-16.59	2	-9.24	2
pgm	267.7	3.70	180704.53	-20.43	1	-19.77	1	-10.15	1
tauD	320.9	3.63	232.83	-17.86	1	-14.41	1	-7.42	1
gadB	123.2	3.55	6068021.53	-23.31	1	-18.97	1	-9.62	1
proP	37.8	3.54	1796.10	-24.67	1	-20.18	1	-12.76	1
dps	75.5	3.53	33.72	-19.87	1	-18.08	1	-8.46	1
relA	103.2	3.42	5552297.06	-23.39	1	-20.72	1	-10.65	1
icd	1776.9	3.41	7019309.68	-21.60	1	-17.59	1	-9.62	1
ydeP	111.3	3.26	213.10	-19.24	1	-16.15	1	-9.38	1
evgA	23	3.22	1.00	-20.91	2	-20.18	1	-12.76	2
gstB	26.4	3.16	292520.98	-19.69	0	-15.89	0	-8.53	0
sdhA	43	3.11	25.64	-19.19	1	-15.89	1	-8.47	2
groL	565.1	2.99	3486871.84	-23.70	1	-19.62	1	-10.00	1
ydjA	2115	2.99	86.30	-18.05	1	-13.51	1	-6.48	2
rspB	15.1	2.93	4021512.13	-24.59	1	-20.77	1	-12.29	1
hemG	23.5	2.92	0.28	-21.07	2	-18.51	2	-10.56	1
yfgM	12	2.90	2418.35	-30.26	1	-25.52	1	-12.76	1
nhaR	18.2	2.87	145487.60	-20.33	1	-17.57	1	-9.62	1
csrA	131.4	2.71							
sdiA	25.7	2.69	366.80	-19.33	1	-17.19	1	-9.71	1
ybeL	221.3	2.62	283.33	-20.00	0	-17.04	0	-9.62	0
ybaL	18.4	2.51	0.03	-18.20	2	-17.33	2	-9.91	2
fdoH	13.3	2.45	0.33	-22.63	2	-18.59	2	-10.62	2
elaB	134.3	2.44	596.72	-19.42	1	-15.90	1	-7.97	1
crp	55.8	2.33	0.96	-24.31	2	-20.18	2	-12.76	2
cysJ	42	2.26	1.00	-17.10	2	-12.76	2	-7.42	2

Table 4.2 (continued)

mRNA Target	Average Free Fluo. (au)	Measured Repression Ratio (No CsrA/CsrA)	Predicted Repression Ratio (No CsrA/CsrA)	Dimer Model ΔG CsrAbind	Dimer Pred	Two Sites ΔG CsrA bind	Two Sites Pred	Single Site ΔG CsrAbind	Single Site Pred
cstA	58.2	2.22	567621.03	-22.57	1	-18.53	1	-10.56	1
fabB	19.3	2.09	1.00	-24.49	2	-19.75	2	-12.76	2
dkgA	2051.3	2.04	139813.34	-24.49	1	-21.22	1	-11.09	1
ppc	182.5	2.02	396.71	-22.92	1	-19.24	1	-9.62	2
pspA	63.2	2.02	59.15	-17.04	1	-12.80	1	-7.97	2
yhiI	18	1.97	31338.45	-19.77	1	-16.77	1	-9.30	2
ahpC	2868.2	1.96	2.05	-19.84	1	-17.59	1	-8.48	2
uxaA	121.3	1.93	0.15	-19.44	2	-18.58	2	-10.61	2
sucB	78	1.90	1023931.20	-24.36	0	-19.62	0	-11.35	0
ppk	35.7	1.88	626.66	-23.93	1	-19.75	1	-10.13	1
sucC	1867	1.81	3.81	-27.08	1	-22.38	1	-12.76	1
yebE	1113.7	1.79	8204.50	-19.43	1	-14.69	1	-8.10	2
truC	10.2	1.67	111.52	-20.49	1	-18.78	2	-9.46	2
yaeP	22.8	1.64	5.94	-20.48	1	-16.80	1	-8.83	2
glcB	336.2	1.64	57.42	-21.03	1	-17.86	1	-9.62	2
dsrB	2365.3	1.63	4474338.32	-23.96	1	-20.28	1	-10.15	2
cysK	305.7	1.61	23231.32	-22.50	1	-18.62	1	-10.65	1
lsrF	966.5	1.61	1.25	-20.55	1	-18.33	1	-9.62	2
ycaC	19.5	1.57	1.00	-18.41	0	-14.69	0	-7.51	0
iscS	321.3	1.55	1.43	-23.94	1	-19.86	1	-10.56	2
yecC	14.3	1.53	91.14	-23.53	1	-19.24	1	-9.62	1
katG	1655.9	1.49	1.00	-16.40	4	-11.90	4	-6.13	4
bfr	74.2	1.43	0.26	-18.33	4	-15.07	4	-7.97	4
csiD	698.2	1.41	0.41	-20.84	4	-16.09	4	-9.23	4
rodZ	23.6	1.40	1.00	-19.69	4	-19.04	4	-10.13	4
pepT	1960.7	1.37	327.80	-20.94	3	-18.49	3	-9.36	0
groS	1578.5	1.33	1.00	-20.84	4	-17.04	4	-9.62	4
iscR	248.6	1.32	151.16	-21.64	3	-17.46	3	-9.49	4
dnaK	859	1.27	1.00	-19.09	4	-17.83	3	-9.62	4
flu	165.9	1.26	9576766	-25.04	0	-20.59	0	-12.27	0
sucA	13.9	1.24	24006.25	-21.75	3	-18.03	3	-9.13	4
aroG	271.6	1.22	0.39	-22.05	4	-17.71	4	-9.62	4
fabI	279.2	1.20	130796.43	-23.01	3	-18.42	3	-8.80	3
carB	54.5	1.16	650.83	-22.92	3	-19.24	3	-9.62	3
yceD	46.9	1.14	1.98	-19.04	3	-17.00	3	-9.62	3

Table 4.2 (continued)

mRNA Target	Average Free Fluo. (au)	Measured Repression Ratio (No CsrA/CsrA)	Predicted Repression Ratio (No CsrA/CsrA)	Dimer Model ΔG CsrAbind	Dimer Pred	Two Sites ΔG CsrA bind	Two Sites Pred	Single Site ΔG CsrAbind	Single Site Pred
pntB	18.4	1.12	207.03	-19.41	3	-15.99	3	-8.02	3
proB	8	1.12	0.26	-18.48	4	-15.94	3	-7.97	4
fecA	44.7	1.10							
hchA	544.9	1.09	99829.30	-24.20	3	-19.46	3	-9.84	3
yjbD	14.8	1.09	0.68	-20.88	4	-19.24	4	-9.62	3
eno	29.7	1.08	1.00	-21.17	4	-16.48	4	-8.00	4
talA	1257.6	1.06	1373178.	-18.29	3	-14.49	4	-6.64	4
mreB	12.3	1.03	0.76	-18.67	4	-14.49	3	-7.85	3
uxaC	13.4	1.03	1982724.	-23.42	3	-18.78	3	-9.62	4
pflB	98	1.03	0.41	-22.76	4	-18.78	4	-10.46	4
ydcS	535.2	1.00	1.00	-23.98	4	-20.73	3	-12.76	4
cdd	80.3	0.99	766.38	-18.54	3	-14.00	3	-7.01	3
yeaH	37.6	0.98	0.96	-18.52	4	-14.54	4	-7.52	3
sdhB	132.9	0.95	233.27	-20.09	3	-15.39	3	-7.97	4
wrbA	81.7	0.94	0.15	-23.15	4	-18.56	4	-9.62	4
pdxB	36.5	0.93	5.58	-17.53	3	-14.31	3	-7.83	3
metC	14.3	0.92	1.00	-17.50	4	-12.96	3	-6.48	4
lpxC	33.8	0.91							
proS	84.6	0.89	426.84	-16.03	3	-11.45	3	-6.16	3
glnS	83.1	0.88	1.00	-20.47	4	-17.59	4	-9.62	4
katE	135.8	0.82	21.60	-22.69	3	-18.81	3	-9.58	3
ompR	11.7	0.81	93.19	-20.50	3	-17.59	3	-9.62	4
guaA	11.4	0.81	188.95	-21.87	3	-17.59	3	-9.62	4
kdsA	129.6	0.80	21552.20	-23.28	3	-20.73	3	-11.11	3
aroD	26.8	0.78	277.42	-20.23	3	-17.58	3	-9.62	3
yqjD	2033.8	0.77	1.88	-21.73	3	-20.31	4	-10.07	3
pta	10.7	0.76	1.86	-18.13	3	-16.08	3	-8.11	3
uxuB	30.5	0.75	0.46	-17.77	4	-16.54	4	-8.61	4
purM	133.1	0.73	1.00	-21.06	4	-18.46	4	-9.62	4
fhuA	9.2	0.73	1.00	-14.18	4	-11.28	4	-6.45	4
cspE	295.7	0.70	1.14	-21.82	4	-19.62	4	-10.00	4
iaaA	42.2	0.70	233025.61	-23.07	3	-18.83	3	-9.62	3
glmS	10.3	0.69	0.70	-22.61	4	-19.09	4	-9.96	4
mtlD	328	0.65	297077.25	-20.28	3	-15.84	3	-7.85	3
acs	1864.2	0.64	9.96	-18.71	3	-14.99	3	-8.51	3

Table 4.2 (continued)

mRNA Target	Average Free Fluo. (au)	Measured Repression Ratio (No CsrA/CsrA)	Predicted Repression Ratio (No CsrA/CsrA)	Dimer Model ΔG CsrAbind	Dimer Pred	Two Sites ΔG CsrA bind	Two Sites Pred	Single Site ΔG CsrAbind	Single Site Pred
manX	17	0.59	115465.61	-24.04	3	-20.69	3	-12.76	3
ackA	7.2	0.16	1905.11	-16.15	3	-11.41	3	-6.99	3

The table contains all active mRNAs tested by the 5'UTR assay and includes the observed fluorescence measurement, the observed ratio of CsrA repression, the predicted ratio of repression from the Dimer model of CsrA interactions, and the predicted ΔG of CsrA binding with the three evaluated model parameters (Single site, Two sites, and Dimer). The numbers in the prediction columns correspond to the following interpretation of the model prediction: 1 – inhibition predicted and observed, 2 – inhibition observed, but not predicted, 3 – inhibition not observed, but predicted, 4 – inhibition neither predicted nor observed, 0 – experimental data failed statistical significance test.

The first observation is that the predicted rates of repression are much greater than what is detectable in the translation assay. This effectively makes the predicted translation rate more of a binary prediction, as either CsrA represses translation or has no effect on translation. It is inappropriate to compare predictions of translation activation to the observations due to fact that the observed activation of targets is likely an indirect effect of CsrA changing cell growth rate (discussed in Chapter 3). For the purposes of evaluating the model, only the inhibition predictions are verifiable.

4.2.3 The Dimer model of CsrA binding is the most accurate prediction of CsrA activity

The first question that needed to be answered was whether the assumptions made for the CsrA binding site predictions are valuable. The three different binding site predictions are 1) CsrA binding to only one site, 2) CsrA binding to two sites, and 3) CsrA binding to two sites with dimerization. If the prediction of CsrA affinity is ignored and the only consideration is the prediction of translation inhibition for the most likely CsrA binding site, then the models behave as summarized in Table 4.3.

Table 4.3: Comparison of observed and predicted inhibition of targets with three different CsrA binding models

	Predicted and observed inhibition	Failed to predict inhibition	Falsely predicts inhibition	No inhibition predicted or observed	Fraction of inhibited genes identified	Fraction of falsely identified genes
Dimer model	65	18	29	23	0.78	0.56
Two sites model	62	21	31	21	0.75	0.60
Single site model	47	36	27	25	0.57	0.52

As can be seen in Table 4.3, the Single site prediction model is capable of identifying 57% of the inhibited targets that we observed as inhibited and identifies 52% of all the targets that were unaffected by CsrA as potential targets. This indicates that the single CsrA binding site prediction fails to predict a large percentage of mRNA targets and performs poorly when identifying targets from non-targets. This shows that a simple, single site model of CsrA-target interactions is inadequate to describing the physical system.

The Two sites model has a slight selective advantage in its overall ability to identify targets and non-targets, but this advantage is smaller than the Dimer model. When the dimerization term is included, the predicted binding sites have a 78% chance of identifying a gene inhibited by CsrA, but will falsely identify 56% of non-targets. While the overall numbers of targets correctly or incorrectly identified as CsrA targets is similar in both the dimer and two site models, the consideration of dimerization does improve the ability of the model to correctly identify mRNA targets and suggests that this model is

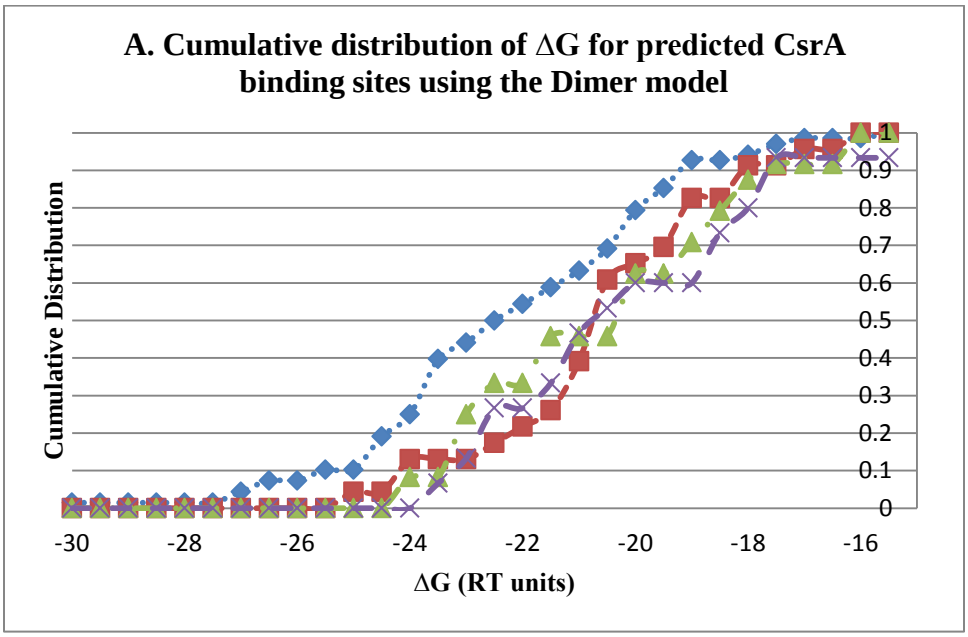
the most complete description of CsrA binding. However, the overall prediction generates too many false positives to be useful to predict mRNA targets of CsrA.

This failure to predict targets based on the translation prediction alone is a direct result of the assumption that CsrA has to bind the mRNA target. In the prediction of CsrA bound translation rate, the model is effectively estimating the translation rate when the concentration of CsrA is infinite. Under this estimation, CsrA would be forced to bind to transcripts at exceptionally weak binding sites that would not be feasible under normal physiological conditions or the conditions tested in the 5'UTR translation assay (Chapter 3.2.1). Given that CsrA binds to a sequence similar to the Shine-Delgarno site, at high enough concentrations, CsrA will likely inhibit most translation initiation sites. For this reason, it is necessary to consider the affinity of the predicted CsrA target when considering the validity of the model.

4.2.4 Population of correctly predicted targets have stronger ΔG of CsrA binding

Given that this model has two distinct parts, one predicting translation and the other predicting CsrA affinity to specific binding sites, we wanted to investigate if the predicted putative strength of the CsrA interaction was related to the specificity of the prediction. To answer this question, the cumulative distribution of the target's CsrA affinity was plotted for all four predictive outcomes in each model variation presented in Table 4.2 (Figure 4.2).

Figure 4.2: Cumulative distribution for tested mRNAs using different models for calculating CsrA affinity



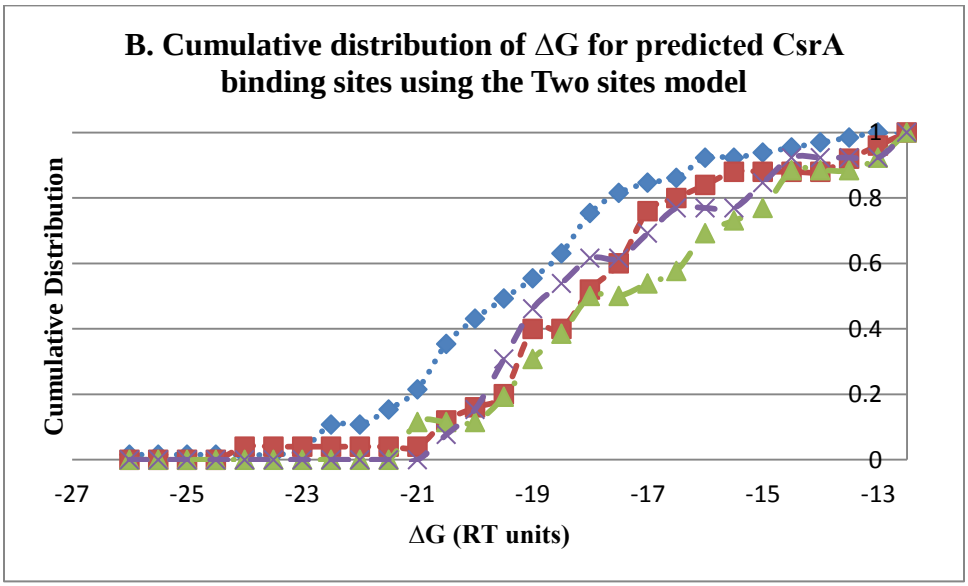
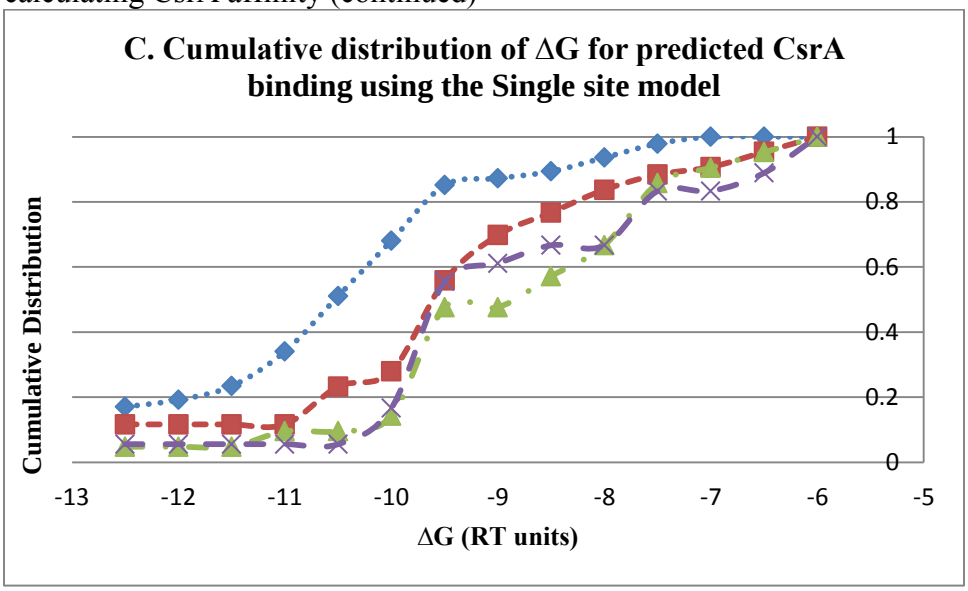


Figure 4.2: Cumulative distribution for tested mRNAs using different models for calculating CsrA affinity (continued)



These are the cumulative distributions of the predicted ΔG of CsrA binding for the four possible predictive outcomes from the three methods to predict CsrA binding.
A) Dimer model
B) Two sites model
C) Single site model
- Blue diamond curves represent the distribution of targets that are positively identified.
- Red boxes are the population where the model failed to predict inhibition.
- Green triangles are the population that was falsely identified as CsrA targets.
- Purple Xs are the population that was correctly identified as having no regulation by CsrA.

As can be seen in the cumulative distribution of positively identified targets, there is a clear differentiation in the distribution of the population of correctly identified targets and incorrectly identified targets at any predicted ΔG of CsrA binding. A comparison in of the average value of ΔG found in the population appears in Table 4.4.

Table 4.4: Average predicted ΔG of CsrA binding for each model

Predicted Population	Dimer Model	Two Site Model	One Site Model	Dimer Model (no motif based predictions)
1	-22.31 ±2.70	-18.83 ±2.42	-10.57 ±1.42	-21.93 ±2.60
2	-20.70 ±2.20	-17.46 ±2.53	-9.52 ±1.61	-20.26 ±1.86
3	-20.62 ±2.46	-16.78 ±2.67	-8.86 ±1.53	-20.52 ±2.46
4	-20.20 ±2.64	-17.31 ±2.58	-9.00 ±1.62	-20.20 ±2.64

Each column contains the average ΔG of CsrA with plus/minus standard deviation (in arbitrary RT units) for the population's predicted outcome of that method. The number corresponding to the Predicted Populations are identical to those described in Table 4.2. The fourth column represents the average populations from the Dimer model while excluding the targets that were identified based on motif predictions.

As can be seen by the average values of the populations, the targets that are predicted accurately (Population #1) have a considerably higher ΔG of CsrA binding than all other populations. This difference is significant with a p-value below 0.05, as determined by the Student's t-test. For a comparison, the Dimer model predicts that randomized segments of RNA of the approximate length of the 5'UTRs tested (200 nucleotides) have an average ΔG of CsrA binding of 19.34±2.18 (RT units). This value was calculated from 100 randomized sequences.

Since some of the targets evaluated were selected based on the similarity of the sequence to the predicted motif, these targets were removed from the average calculated

in column 4. While the removal of these targets does lower the overall ΔG of CsrA binding, there is still a bias for stronger binding in the accurately predicted targets over the falsely identified targets. By excluding these targets from the population, it is clear that this bias in binding free energy was not artificially introduced by our model assumptions. The bias for higher predicted binding in the true target clearly demonstrates that the model is identifying the right parameters to consider when predicting CsrA binding.

4.2.5 Accurate predictions of mRNA targets using ΔG reveals the significance of the position weight matrix

For this model to be applied as a predictive method for identifying CsrA regulated genes, it is necessary to consider when the free energy of CsrA binding will maximize correct target predictions and minimize incorrect targets predictions. For the Dimer model, this maximum difference between the percent of true positive and false positive predictions occurs at -23.5 (RT units), as seen in Figure 4.2. At this cut off free energy, the Dimer model selects 42% of the population of the true positives and 13% of the false positive targets. The different models were evaluated at the following predicted free energy of CsrA binding: Dimer model - -23.5 (RT units), Two sites model - -20.0 (RT units), Single site model - -10.5 (RT units).

When these free energy considerations are applied to targets tested in our translation assay, we noticed that many of the targets identified are identical with respect to the model. For example, comparing the Dimer model and the Single site model, 26 mRNA targets are identified by both models, 11 targets are predicted exclusively by the dimer model, and 5 targets are exclusive predictions of the Single site model.

Interestingly, the Single site model performed almost as well the Dimer model when it came to the most accurate prediction of CsrA targets. This can be explained when

one considers the free energy model used to make this prediction. The single site model has a minimum free energy prediction of -12.76 (RT units) when all nucleotides match the PWM (Table 4.1). A free energy prediction of -10.5 in the single site model means that only three possible CsrA binding sites could be considered, AAGGA, AAGGN, or ANGGA. This suggests that having a strong CsrA binding site in a position that affects ribosome binding is a very strong indicator of a gene being a target of CsrA.

4.2.6 Prediction of CsrA affinity has qualitative significance

Understanding how CsrA is capable of regulating so many cellular factors is critical towards fully understanding the complexity of global CsrA-mediated regulation. One hypothesis is that CsrA displays differential affinities for the targets it regulates allowing for CsrA to be specifically titrated to affect some genes but not others. While the previous section demonstrated that the predicted affinity of the CsrA interaction had a relationship to correct target predictions, we wanted to evaluate if the predicted affinity also had a physiological significance that could explain global CsrA regulation.

To assess the physiological relevance of the free energy predictions, we devised a means of altering CsrA expression that would alter the fluorescence of the 5'UTR translation assay. The rationale for this assay is as follows: 1) Fluorescence from a specific mRNA will display one activity when CsrA is bound and a different activity when CsrA is not bound. 2) The ratio of bound and unbound mRNA will follow a standard sigmoidal relationship dependent on the concentration of CsrA, the mRNA, and their affinity towards each other. 3) Observable fluorescence will reflect the ratio of the

bound and unbound mRNA transcripts. Using this logic, one would expect that population's fluorescence, with a specific mRNA, will resemble a sigmoidal binding curve dependent on the amount of free CsrA present and the affinity that CsrA has to the mRNA. To test this system, CsrA expression was titrated with *glgC* (a high affinity prediction) and *aidB* (a low affinity prediction).

For this data, the fluorescence ratio is representative of the ratio of bound to unbound mRNA transcript with 0% bound mRNA being defined as repression ratio of 1 and 100% bound mRNA having a value between 8-16. This value is dependent on the rate of cell division after induction of CsrA (see Methods 4.4.1 for more details).

Expressing CsrA across a thousand fold range approximately revealed the characteristic sigmoidal increase in repression level for both *glgC* and *aidB* (Figure 4.3). For comparison purposes, an approximated sigmoidal binding curve is drawn to fit the data. Specifically, the equivalence point for *aidB* is shifted towards higher expression levels of CsrA with respect to *glgC*, suggesting that *glgC* is more sensitive (has a higher affinity) to CsrA than *aidB*. This is evidence that CsrA is capable of preferentially recognizing targets at specified concentrations and that our estimations for CsrA affinity to a target have physiological significance relative to each other.

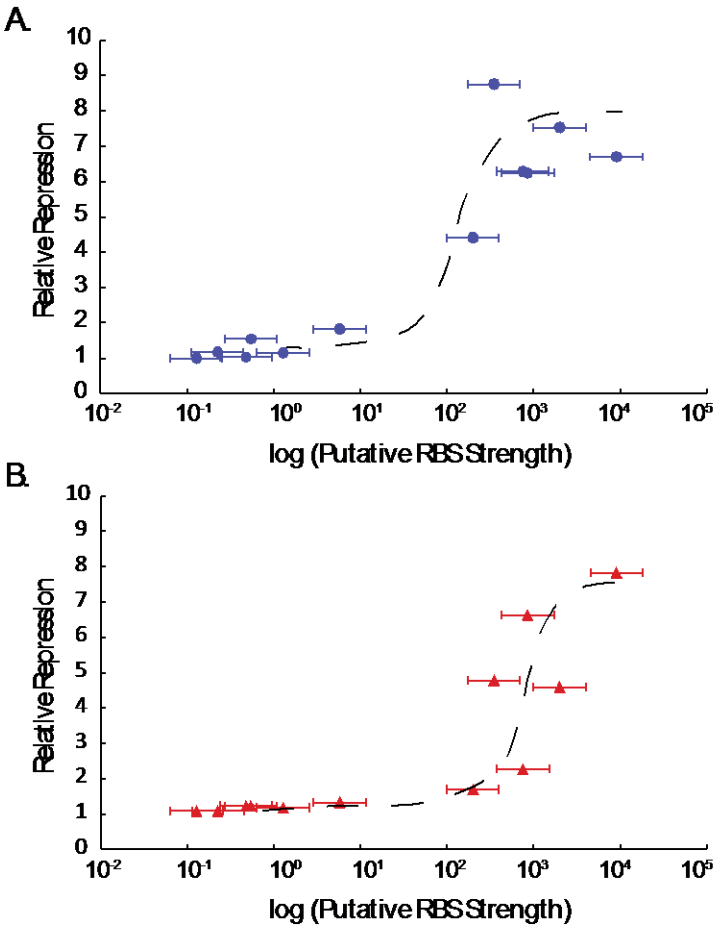


Figure 4.3: Relative repression ratios of select targets in response to alterations of CsrA expression

(A) Expression profile of *glgC* mRNA in 5'UTR translation assay

(B) Expression profile of *aidB* mRNA in 5'UTR translation assay

4.3 DISCUSSION

This chapter presents a model devised from first principles of CsrA binding and translation initiation to predict the free energy of potential CsrA interactions and the effects that the interactions will have on a specific gene. While each of the three models over estimates the total number of mRNAs affected by CsrA, consideration of the predicted strength of the CsrA interaction allows for predictions that reduce the expected number of false positives. Currently, the Dimer model is being applied to the *E. coli* genome and will be used to predict novel targets of CsrA. A more complete knowledge of the CsrA regulated genes will greatly improve a current hole in the understanding of cellular regulation, allowing for a more integrated understanding of the interplay between transcriptional-translational regulation.

Furthermore, the comparison of the different models to predict CsrA binding demonstrates the missing information in the mechanistic understanding of CsrA based regulation. While the models demonstrate that the presence of an AAGGA sequence near the RBS site was a strong indicator of CsrA based regulation, it is far from a presenting a complete understanding of CsrA based prediction. This work emphasizes the diversity of CsrA based interactions. While many genes analyzed showed CsrA binding patterns that correlate the current understanding of CsrA-RNA interactions, many genes did not.

In particular, *thiG* mRNA sequence had a very low predicted affinity for CsrA but displayed very strong regulation by CsrA in the translational assay. Genes that significantly deviate from the predicted interaction patterns represent interesting test subjects that can be analyzed to better the overall understanding of CsrA-RNA interactions and improve the predictive binding model. A more thorough characterization of this mRNA's interaction with CsrA would be in order. At the very least, this

information should inform the design of synthetic RBS sequences to avoid containing an AAGGA sequence to prevent unintentional regulation.

Additionally, the models presented provide a framework to explain the overall strategy that the cell employs when using CsrA to regulate entire pathways. The implication that CsrA is specifically titrated to unique genes suggests that CsrA based regulation is more nuanced than a binary regulation-no regulation understanding of CsrA-mRNA interactions. This finding is reinforced by recently published work that has shown that the binding of RsmE (CsrA) to its sRNA follows a precise sequence and is based on the concentration of the available mRNA (108). This information that the sRNA competitor of a CsrA like protein has specific titratable levels of protein binding strongly suggests that the availability of free protein is specifically manipulated to allow the regulation of certain mRNAs before regulating others. We hypothesize that genes that produce proteins related to specific tasks will have similar affinities to CsrA, thereby allowing CsrA to have a hierarchical control over specific subsets of genes. This hypothesis will be tested as part of the ongoing work to apply this model to the genome of *E. coli.*

4.4 METHODS

4.4.1 Evaluation of CsrA affinity through titration of fluorescence signal

Titration of CsrA concentration was achieved through modification of the CsrA expression system employed in the 5' UTR assay discussed in Chapter 3. Eight RBS sites were created to alter the expression of CsrA from the vector used in the mentioned assay. The RBS sequences were calculated by the RBS Calculator to express the protein across a thousand-fold range of the original system. The eight sites have computationally-determined putative levels of CsrA expression of 9000, 2000, 1000, 850, 750, 500, 350,

and 200 (with 9000X being the predicted rate of expression from the original plasmid). These sites were created on primers that amplified CsrA so that they could be ligated in to the plasmid using standard techniques. A list of the 8 RBS sequences and the primers used to attach them to CsrA appear in Appendix Table A.1.

Using these modifications to the CsrA expression system, the 5'UTR-translation assay was carried out for each CsrA expression modification. In these tests, the ratio of repression is measured as in Chapter 3 with the exception that all fluorescence values were normalized to the highest fluorescence measured. In this case, that value corresponded to the fluorescence measured when the weakest RBS (200) was uninduced. The maximum ratio of fluorescence from the uninduced and the induced CsrA conditions signals 100% CsrA-mRNA binding. This ratio is related to the number of cell divisions that occur to allow depletion of the fluorescence signal (as described in Chapter 3). In this case, the maximum value is between 8 and 16.

For these experiments, the activity of the RBS (and overall concentration of CsrA) should be related to the putative strength of the RBS and be proportional to the concentration of the mRNA for CsrA. For induced and uninduced measurements, the amount of transcript present was viewed as 100% available following induction and 1% available for uninduced cells. These estimates for the leakiness of the CsrA promoter come from previously published characterization of the system (93). Additionally, the putative RBS strength is expected to have a potential error of 2X the putative value (between 50% and 200%).

Chapter 5

Conclusion and Future Directions

The work presented in this dissertation represents the development of unique applications and techniques for the study of RNA-protein interactions and resulted in the advancement of the current understanding of CsrA-RNA interactions.

Chapter 2: Development of a Tri-molecular Fluorescent Complementation assay (TriFC) to detect regulatory RNA-protein interactions

Through the application of tri-molecular fluorescence complementation (TriFC) and some basic protein and RNA design principles, we have demonstrated how it is possible to visualize RNA-protein interactions *in vivo* and have demonstrated that the method is sensitive enough to detect mutations that alter the affinity of the interaction. As this tool represents a method to detect the interaction of CsrA and CsrB, it would be extremely useful to adapt this tool to so that the modified CsrB molecule was expressed under the natural system. If this could be done with TriFC, one could imagine that it would be possible to subject the cell to extracellular stresses to understand the conditions that activate CsrB.

Applications of the TriFC method could be extended to probing RNA-protein interfaces. The TriFC fluorescence output was shown to be sensitive to mutations in CsrA that affect the affinity of the RNA protein interaction. Using mutational libraries of the target RNA or protein, it would be possible to identify mutations that negatively affect the RNA-protein interaction that could be used to elucidate the structure of the RNA- protein interaction.

Chapter 3: Translational assay for characterizing CsrA regulation on mRNA transcripts with interaction validation using TriFC

This work presented one of the first largescale characterizations of mRNA-CsrA interactions through the application of a 5'UTR translational fusion. In this assay, the 5'UTR and a section of the coding sequence of an mRNA of interest was expressed as a translational fusion to GFP. The fluorescence of the system is directly related to the translational activity of the mRNA being tested. If the mRNA was a target of CsrA, induction of CsrA alters the translation activity of the target. This work is one of the first attempts at quantifying the effect of CsrA on mRNA activity and has directly introduced evidence that mRNA-CsrA interactions exist beyond a simple binary output of regulation vs no regulation. Indeed, this work has demonstrated that specific sequences of RNA show gradations of responses to CsrA.

The concept that separate RNAs can have a gradient of responses to CsrA should be pursued further. The evidence suggests that CsrA does not totally inhibit the mRNA targets that it interacts with. One possible explanation for this observation is that CsrA has low affinity for the mRNA target and does not completely bind all the available mRNA. This explanation could explain how CsrA can differentially affect so many mRNA targets. Another explanation for incomplete inhibition could be that CsrA binding to an mRNA target causes a partial shift in its translational activity. This would be a unique mechanism to biological systems, which tend to be simplified as either inhibiting vs non-inhibiting. If this interaction can be verified, it may change the current understanding of cellular regulation.

Furthermore, the TriFC tool was applied to verify mRNA-CsrA target interactions. With some relatively minor changes to the design of the RNA target molecule that directs TriFC, we were able to create a system that allowed for *in vivo* screening of mRNA-CsrA interactions. This work provided the first direct evidence for mRNA-CsrA interactions of 19 targets, including the sigma factor *rpoS*. While this work

expands the current knowledge of mRNA targets of CsrA, the knowledge of CsrA regulation is far from complete. Further characterization of the CsrA network should continue to inform and improve models of regulation.

Chapter 4: A first principles model of CsrA binding and activity with experimental validation

A model was created using the primary principles of RNA-CsrA interactions to predict RNA-CsrA interaction affinity and to predict the effect that CsrA would have on translation initiation. The accuracy of this model was compared to experimental data gathered from the 5'UTR translation assay presented in Chapter 3. Comparison of the model predictions to the translational assay data clearly demonstrates that the current understanding of mRNA-CsrA interactions is adequate to describe interaction behavior and has predictive power to assess the likelihood that a given sequence may be a target of CsrA. However, the data presented shows that certain sequences represent a clear deviation from the accepted understanding of mRNA-CsrA interactions. Further study of the genes that didn't comply with current model could greatly improve what is known of CsrA-mRNA interactions.

Evaluation of this model's prediction of mRNA-CsrA interaction suggested that mRNA sequences are uniquely titrated by the concentration of CsrA to allow finely tuned response of targets based on the availability of CsrA. This model and data from the 5' translation assay strongly supports the current hypothesis that mRNA-CsrA affinity is the main mechanism for CsrA fine tuning of gene expression. Recent work investigating the relationship between an analogue of CsrA and its regulatory RNA has shown that the RNA-protein interaction is highly tuned and ordered as to allow protein binding at specific locations only when the protein concentration is high enough (108). This information taken with the findings that specific mRNAs have a concentration and

sequence dependent response to CsrA is strong evidence that CsrA availability can be tuned to affect specific targets and pathways. Further experimentation with CsrA should strive to demonstrate that the cell employs condition specific titration of CsrA concentration to target the mRNAs needed for specific responses.

This method of genetic control could also be implemented in engineered pathways. Using the concepts of the RNA binding molecules and the model presented to predict CsrA-mRNA interactions, it should be possible to design novel mRNA sequences designed to work with orthologous RNA binding proteins to give tunable post-transcriptional control of translation initiation.

Appendix

Table A.1 mRNA sequences tested in translational assay and TriFC assay

Gene name	5'UTR-100CDS
glgC	TCTGGCAGGGACCTGCACACGGATTGTGTGTGTTCCAGAGATGATAAAAAAGGAGTTAGTCATGGTTAGTTTAGAGAAGAACGATCA CTTAATGTTGGCGCGCCAGCTGCCATTGAAATCTGTTGCCCTGATACTGGCGGGAGGACGTGGTACCCGCCTGA
pckA	CGTTTCGTGACAGGAATCACGGAGTTTTTTGTCAAATATGAATTTCTCCAGATACGTAAATCTATGAGCCTTGTCGCGGTTAACACCCC CAAAAAGACTTTACTATTCAGGCAATACATATTGGCTAAGGAGCAGTGAAATGCGCGTTAACAATGGTTTGACCCCGCAAGAACTCGA GGCTTATGGTATCAGTGACGTACATGATATCGTTTACAACCCAAGCTACGACCTGCTGTATC
glsA	ATTCGCGCATTGAGCGAGTGGCTGAATGAAAATACCGATCTTGATGTGACCTTTATTGATATTCCTAATCCTGCATAACGAATAATCAG AGGGATCGAAAGTGCAACGAGCGCGTTGTTATCTGATAGGTGAAACGGCGGTAGTGCTGGAACTGGAACCGCCGGTGACGCTGGCT AGCCAGAAACGGATCTGGCGACTGG
aidB	GTGACTGCCATTGATGGAGGGAGACACAgtgCACTGGCAAACTCACACCGTTTTTAATCAACCTATACCATTAAATAACAGCAACTTAT ACCTGTCTGATGGCGCGCTCTGCGAAGCGGTAACGCGTG
rseA	GAGACAGATAGTTTTCCGAACTATTGAGTCCCTCCCGGAAGATTTACGCATGGCAATAACCTTGCGGGAGCTGGATGGCCTGAGCTAT GAAGAGATAGCCGCTATCATGGATTGTCCGGTAGGTACGGTGCGTTCACGTATCTTCCGAGCGAGGGAAGCTATTGATAACAAAGTT CAACCGCTTATCAGGCGTTGACGATAGCGGGATACTGGATAAGGGTATTAGGCATGCAGAAAGAACAACTTTCCGCTTTAATGGATG GCGAAACGCTGGATAGTGAGCTGCTTAACGAACTGGCTCATAACCCAGAAATGCAGAAAACCTGGG
maeB	GTCTGTAGACTCCGGCAGGGTAATAATGTGCGCCACGTTGTGGGCAGGGGCTGTTGCCCACACACTTTATTTGTGAACGTTACGTGAA AGGAACAACCAAATGGATGACCAGTTAAAACAAAGTGCACTTGATTTCCATGAATTTCCAGTTCCAGGGAAAATCCAGGTTTCTCCAA CCAAGCCTCTGGCAACACAGCGCG
thiG	GCGATTAATCAGCAAATCGTCCCGCGTGAGCAGTGGGCGCAACATATCGTGCAGGATGGCGACCAGATCCTGCTTTTTCAGGTTATTG CAGGGGGTTGAAATGTTACGTATTGCGGACAAAACGTTTGATTCACATCTGTTTACCGGCACAGGGAAATTCGCTTCTTCACAACTGA TGGTGGAGGCGATCCGCGCTTCCG
yidQ	GCGATAGCCGCCACAGTGCGACGTAATGATGATGTTATTTTCAGCAAGGGATCGCTATAACTCATCAGGCCAACACCCACGATCAACA GATGAATTTATTATGATAAGAAATGTGTTGTTAGCGTTCATGATATGCAGCGGAATGACATTACTCGGAGGATGCTCCAGCGTGATGT CCCATACCGGAGGGAAAGAAGGAA
yafQ	ATCAATTAACCATTCAATCAATCAAAAACAGCGAAGCTGGCATTGATGTTCATAAGGCCAAAGACGCCGATGATTTATTTGATAAATTA GGAATTTAAATATGATTCAAAGGGATATTGAATACTCGGGACAATATTCAAAGGATGTAAAACTTGCACAAAAGCGTCATAAGGATAT GAATAAATTGAAATATCTTATGA
mscS	ACTGCCCGTCATAAGGCGGCGGAGTGTATTTCTCCATTTTGAGTCAGTTGAAAAGGAATATTGAATGGAAGATTTGAATGTTGTCGAT AGCATAAACGGCGCGGGAAGCTGGCTGGTAGCTAACCAGGCGCTGCTGCTAAGTTATGCAGTAAACATCGTGGCGG
uspG	GAATGATTTGTTCATGATTAACAGGGAGAAAGGGTTATGTATAAGACAATCATTATGCCAGTTGATGTATTTGAAATGGAATTGAGCG ACAAAGCTGTTCGCCACGCTGAATTCCTCGCCCAGGATGACGGAGTTA
deoD	ATTTTGGTACTTCTGATATGGAATATGGCAAAGCCATGTTCTGATGGATTTGGGCGGAGCGTTGACTCCGCCTTTGTTATGTCACAAAA AGGATAAAACAATGGCTACCCCACACATTAATGCAGAAATGGGCGATTTCGCTGACGTAGTTTTGATGCCAGGCGACCCGCTGCGTG CGAAGTATATTGCTGAAACTTTCC
uidR	TTCACGCTAACTATAATGGTTTAATGATGGATAACATGCAGACTGAAGCACAACCGACACGGACCCGGATCCTCAATGCTGCCAGAGA GATTTTTTCAGAAAATGGATTTCACAGTGCCTCGA
acnA	TCACCCTGAAGAGAATCAGGGCTTCGCAACCCTGTCATTAAGGAGGAGCTatgTCGTCAACCCTACGAGAAGCCAGTAAGGACACGTT GCAGGCCAAAGATAAAACTTACCACTACTACAGCCTGCCGCTTGCTGCTAAATCACTGGGCG
ydhQ	GTCGTGGATAAATCCTCCGTAGCCTTTACACTTAGGTAGCAATAAAGGCACACAGATGCAGCAAAGGAATAAATCATGGGATCTGAT GCGAAAAACTTGATGAGCGACGGGAACGTGCAAATTGTTAAGACCGGCGAGGTCATTGGCGCGACGCAACTTACTGAAGGCGAGTT GA
frdB	GACGTGAAGATTACTACGCTGCCGCCAGCTAAACGCGTTTACGGTGGCGAAGCGGATGCAGCCGATAAGGCGGAAGCAGCCAATAA GAAGGAGAAGGCGAATGGCTGAGATGAAAAACCTGAAAATTGAGGTGGTGCGCTATAACCCGGAAGTCGATACCGCACCGCATAGC GCATTCTATGAAGTGCCTTATGACGCAA
Rnk	GAGTTAAACCCTCGCCGCCTGACGGTGAGGGTTTTCTTTTGGGTATGTTTATCTGCGACACTCGCAGCACCGACGACATGGAGTAAAA ATGTCCAGACCAACTATCATCATTAACGACCTGGATGCCGAACGCATCGATATTCTGCTGGAGCAACCCGCCTATGCTGGTTTGCCAAT CGCCGACGCGT
hflK	GATTACCTGATCTAACGGCGTAGCGTCTGAAGCGTGGAGTCATATCCTCTGGCGTCGAAAGACAACAGGGATCACCGCATAACAAAT ATGGAGCACAAACATGGCGTGGAATCAGCCCGGTAATAACGGACAAGACCGCGACCCGTGGGGAAGCAGCAAACCTGGCGGCAACT CTGAGGGAAATGGAAACAAAGGCGGTC
astD	CCGAGCGTTTGATTTTAACCGCCGCACAACTGGATGCCCTCAAATGCCACGCCGGGGATCGCGTTCGTCTGGTGCGCCTGTGCGCAGA GGAGAAAACAGCATGACTTTATGGATTAACGGTGACTGGATAACGGGCCAGGGCGCATCGCGTGTGAAGCGTAATCCGGTATCGGG CGAGGTGTTATGGCAAGGCAATGATG
poxB	GGCTATTTAACCGTTAGTGCCTCCTTTCTCTCCCATCCCTTCCCCCTCCGTCAGATGAACTAAACTTGTTACCGTTATCACATTCAGGAGA TGGAGAACCatgAAACAAACGGTTGCAGCTTATATCGCCAAAACACTCGAATCGGCAGGGGTGAAACGCATCTGGGGAGTCACAGGC

	GACTCTCTGAACGGTCTTAGTG
nnr	CGCTAACTGATTGAGATCGAGGGGCTCTGACATGACGGACCATACAATGAAGAAAAACCCCGTAAGTATACCACACACCGTCTGGTA CGCCGACGATATCCGCCGCGGAGAACGCGAGGCGGCAGATGTGC
gshB	CCTGTTGGGACCTCGCGTTTTGCGTACAGAGACAACTGCGCTCACCGCCATTACCGCGCTACAAGTACGATTTGGCGATTTGGGCTAA CGGAGAAGAATAATGATCAAGCTCGGCATCGTGATGGACCCCATCGCAAACATCAACATCAAGAAAGATTCCAGTTTTGCTATGTTGC TGGAAGCACAGCGTCGTGGTTACG
fucO	AGTGCTGAGCGATGAAGAGATTGCCGTAGTGCTGGAGAAATTCAAAACCTATGGGTTACGAATTGAAGAGTAATTTCGTAAAGCAAC AAGGAGAAGGATGATGGCTAACAGAATGATTCTGAACGAAACGGCATGGTTTGGTCGGGGTGCTGTTGGGGCTTTAACCGATGAGG TGAAACGCCGTGGTTATCAGAAGGCGC
moaB	GCGGGAGAAGAAACAGACCCATTTCCTGCATCAAAACAACACCGGTATTACGCAAAACTTATCGTACATTGGCGGCTAAAACGTCAAA AGGAGAGATCAGatgAGTCAGGTAAGCACTGAATTTATCCCGACCCGTATTGCTATTCTTACGGTTTCTAATCGTCGCGGTGAAGAAGA CGATACCTCCGGTCACTATCTGC
gstA	AGGGAAACTCTTTTACAAGCCGCTAACTTTTCGTTAGCGGCTTTTTTTTTGTTCAGCAACGCACTACCATACTTTAAACCACAGACAAAA GGAGTTACCGATGAAATTGTTCTACAAACCGGGTGCCTGCTCTCTCGCTTCCCATATCACCCTGCGTGAGAGCGGAAAGGATTTTACCC TCGTCAGTGTGGATTTAATGA
ucpA	AATGGAGTGTAACGCTCTGTATTAACAAGGAGAGCATTAAAATGGGTAAACTCACGGGCAAGACAGCACTGATTACGGGCGCATTGC AGGGAATTGGCGAAGGAATTGCCAGAACTTTTGCACGTCATGGCGCGAACCTAA
cysD	CTGATGGCAAGAAAATAGCGGTATTGCAAAGGAACGGTTATGGATCAAATACGACTTACTCACCTGCGGCAACTGGAGGCGGAAAG CATCCACATTATTCGCGAGGTGGCGGCAGAATTCTCAAATCCGGTGATGCTCT
cmk	TTATGTTAACGGTACGCCTGTTTTAAGGAGATAAAGATGACGGCAATTGCCCCGGTTATTACCATTGATGGCCCAAGCGGTGCAGGGA AAGGCACCTTGTGTAAGGCTATGGCGGAAGCGTTGCAATGGCATCTGC
rnr	ACACGCTTGCCGATTTGGTTGAAGAGAATCAACCGCTTTATAAATTATTGCTGGTGGAGTGACGAAAATCTTCATCAGAGATGACAAC GGAGGAACCGAGATGTCACAAGATCCTTTCCAGGAACGCGAAGCTGAAAAATACGCGAATCCCATCCCTAGTCGGGAATTTATCCTC GAACATTTAACCAAACGTGAAAAAC
hemX	CGCCCGGGAACTGGGCTGGCAAGACATTAAGGTCGCCGATAACGCTGACAACGATGCGCTTTTACGGGCATTACAATAACTCTCATAA CAGGAAGCCATAATGACGGAACAAGAAAAAACCTCCGCCGTGGTTGAAGAGACCAGGGAGGCCGTGGACACCACGTCACAACCTGT CGCAACAGAAAAAAAGAGTAAGAACA
ltaE	ATTTCCCGTCATAATAAGGACATGCCATGATTGATTTACGCAGTGATACCGTTACCCGACCAAGCCGCGCCATGCTCGAAGCGATGAT GGCCGCCCCGGTTGGGGACGACGTTTACGGAGACGACC
glpR	GCGCTGATCTGGATTGTCGCCGGATGGTTTGATTTGTTTGGGATGTCGATGGCGAACGGAGCACACATCGCCGGGTTAGCCGTGGGT TTAGCGATGGCTTTTGTTGATTCGCTCAATGCGCGAAAACGAAAATAATTCCAGGGATTTATAAATGAAACAAACACAACGTCACAAC GGTATTATCGAACTGGTTAAACAGCAGGGTTATGTCAGTACCGAAGAGCTGGTAGAGCATTTCTCCGTCAGCCCGC
entF	TTGTGTGTCAGCCGCAGTCACAGGCGTCCTGCCAGCAGTGGCTGGAAGCCCACTGGCGTACTCTGACACCGACGAATTTTACCCAGTT GCAGGAGGCACAATGAGCCAGCATTTACCTTTGGTCGCCGCACAGCCCGGCATCTGGATGGCAGAAAAACTGTCAGAATTACCCTCC GCCTGGAGCGTGGCGCATTACGTTG
gadA	CCTTCAAATAAATTTAAGGAGTTCGAAATGGACCAGAAGCTGTTAACGGATTTCCGCTCAGAACTACTCGATTCACGTTTTGGCGCAA AGGCCATTTCTACTATCGCGGAGTCAAAACGATTTCCGC
ahr	AACAACACCAGAGAAGGACCAAAAAATGTCGATGATAAAAAGCTATGCCGCAAAAGAAGCGGGCGGCGAACTGGAAGTTTATGAGT ACGATCCCGGTGAGCTGAGGCCACAAGATGTTGAAGTGC
tnaA	TAAATGATGGTGCTTGCATATATATCTGGCGAATTAATCGGTATAGCAGATGTAATATTCACAGGGATCACTGTAATTAAAATAAATG AAGGATTATGTAATGGAAAACTTTAAACATCTCCCTGAACCGTTCCGCATTCGTGTTATTGAGCCAGTAAAACGTACCACTCGCGCTTA TCGTGAAGAGGCAATTATTAAAT
pgaA	AGGCATTGGGATTTATGCCGTATTCCTGAAGATCCTCATCATTGGAATGGATTTTCGGGCGAGAAAAGGATTTTATATGGACACTCTG CTCATCATTTCTTCTTCTCATCATCAACAATTCACGTCTCTCTTCCGCGTTTAATAACGGATTATGAGGTGCAAAAATATCTTTCTTTTCA GTTACCTGTAATTAGATACAGAGAGAGATTTTGGCAATACATGGAGTAATACAGGATGTATTCAAGTAGCAGAAAAAGGTGCCCGAA AACCAAATGGGCTTTGAAACTTCTTACTGCCGCATTTTTAGCAGCGAGTCCCGCGGCGAAGAGTGCTG
hfq	GTATCGTGCGCAATTTTTTCAGAATCGAAAGGTTCAAAGTACAAATAAGCATATAAGGAAAAGAGAGAATGGCTAAGGGGCAATCTT TACAAGATCCGTTCCTGAACGCACTGCGTCGGGAACGTGTTCCAGTTTCTATTTATTTGGTGAATGGTATTAAGCTGCAAG
entC	GAAAATATAAATGATAATCATTATTAAAGCCTTTATCATTTTGTGGAGGATGATATGGATACGTCACTGGCTGAGGAAGTACAGCAGA CCATGGCAACACTTGCGCCCAATCGCTTTTTCTTTATGTCGCCGTACCGCAGTTTTACGACGTCAG
clpB	AGTAGCAACTTTGATCCGTTATGGGAGGAGTTATGCGTCTGGATCGTCTTACTAATAAATTCCAGCTTGCTCTTGCCGATGCCCAATCA CTTGCACTCGGGCACGACAACCAATTTATCGAACCACTTCATT
rpoS	GGAAATCCGTAAACCCGCTGCGTTATTTGCCGCAGCGATAAATCGGCGGAACCAGGCTTTTGCTTGAATGTTCCGTCAAGGGATCACG GGTAGGAGCCACCTTatgAGTCAGAATACGCTGAAAGTTCATGATTTAAATGAAGATGCGGAATTTGATGAGAACGGAGTTGAGGTTT TTGACGAAAAGGCCTTAGTAGAACAGG
uxaB	ATTCAAACAGGTTGTATGACTAATCAGAAGGGAACCCATTGTGAAAACACTAAATCGTCGCGATTTTCCCGGTGCACAGTATCCAGAA CGTATCATTCAGTTTGGTGAAGGTAACTTCCTGCGCGCCTTTGTTGACTGGC
pgm	AGCATGATCGAACACATCATGCAGTCGATGCAATTCCCGGCGGAATTGATTGAGAAGGTTTGCGGAACTATCTAAAACGTTGCAGAC AAAGGACAAAGCAatgGCAATCCACAATCGTGCAGGCCAACCTGCACAACAGAGTGATTTGATTAACGTCGCCCAACTGACGGCGCAA TATTATGTACTGAAACCAGAAGCAG

tauD	TGCTGGCGGGGATCGCGGTGATTGCGATTATCGCCTTTCTTTTAGAACTGGGTCTGCGCGCGTTACAGCGCCGCCTGACGCCCTGGCA TGGAGAAGTACAATGAGTGAACGTCTGAGCATTACCCCGCTGGGGCCGTATATCGGCGCACAAATTTCGGGTGCCGACCTGACGCGC CCGTTAAGCGATAATCAGTTTGAAC
gadB	CGATCCAATCATTTTAAGGAGTTTAAAatgGATAAGAAGCAAGTAACGGATTTAAGGTCGGAACTACTCGATTCACGTTTTGGTGCGAA GTCTATTTCCACTATCGCAGAATCAAAACGTTTTCCGC
proP	GTCATTAACTGCCCAATTCAGGCGTCAACTGGTTTGATTGTACATTCCTTAACCGGAGGGTGTAAGCAAACCCGCTACGCTTGTTACAG AGATTGCATCCTGCAATTCCCGCTCCCCTTTTGCGGCCGTCGCGCTGATTTTTCTGGCGTTTGCGGAAATGGGCCAACTCTGCGAGGAA AGCTATGCTGAAAAGGAAAAAAGTAAAACCGATTACCCTTCGTGATGTCACCATTATTGATGACGGTAAACTGCGTAAAGCCATTACC GCAGCATCACTGGGTA
Dps	GTTAATTACTGGGACATAACATCAAGAGGATATGAAATTATGAGTACCGCTAAATTAGTTAAATCAAAAGCGACCAATCTGCTTTATAC CCGCAACGATGTCTCCGACAGCGAGAAAAAAGCAACAGTAGAGTTGCTGA
relA	CTGCAACGCTGGCTCGGGATAGCGAAGCGTTATTAAAAGCAGGATATACCATTGCGCGACTGGCGATGCTGGATATGTTCCCACACA CGGGACATCTGGAATCGATGGTACTTTTCTCGCGCGTTAAATAGTTGCGATTTGCCGATTTCGGCAGGTCTGGTCCCTAAAGGAGAGG ACGATGGTTGCGGTAAGAAGTGCACATATCAATAAGGCTGGTGAATTTGATCCGGAAAAATGGATCGCAAGTCTGGGTATTACCAGC CAGAAGTCGTGTGAGT
Icd	ATAACGCGCATCTTTCATGACGGCAAACAATAGGGTAGTATTGACAAGCCAATTACAAATCATTAACAAAAAATTGCTCTAAAGCATC CGTATCGCAGGACGCAAACGCATATGCAACGTGGTGGCAGACGAGCAAACCAGTAGCGCTCGAAGGAGAGGTGAATGGAAAGTAA AGTAGTTGTTCCGGCACAAGGCAAGAAGATCACCCTGCAAAACGGCAAACTCAACGTTCCTGAAAATCCGATTATCCCTTACATTGAAA G
ydeP	ATGAGAGATGATGCTTAACTTATCGTATTTAATCTATTGTTTAACGATAATAGATTAAAAATAGAAAGAGAGTAGTCCCCTACTCTTAT ATATCCATGTTGGCGATAATCCCTTTGTATGTACTGTGCATCATCGCTATTACAAATCCTAATAATTCATTTCCACACAGGATAAGTAGA TGAAGAAAAAAATTGAATCCTACCAGGGTGCTGCAGGTGGTTGGGGTGCTGTTAAATCCGTAGCGAATGCAGTACGTAAGCAGATG GATATACGCCAGG
evgA	TACTTGTCGAATTATCTTAAAGGAAGCTCAGATTTTCTTATTTTTATTGAGAAAATGAGATGACGCCTTATGTCTGTATTACTACAGGGA GAAGGGAGATGCTTCATTGCAAAGGGAATAATCTATGAACGCAATAATTATTGATGACCATCCTCTTGCTATCGCAGCAATTCGTAATT TATTGATCAAAAACGATATTGAAATCTTAGCAGAGTTGACTGAAG
gstB	CTGTTCAGCCTGTGGCTGCCTGTCGGAAAAGGGGTGAGTATGATTACGCTGTGGGGTCGGAATAATTCAACTAACGTAAAAAAAGTA TTGCTGACGCTCGAAGAACTGGAACTACCTTATGAGCAAATTCTCGCGGGCC
sdhA	ACGTTAAACCGCTGGCTTTGCGCCTGATGCTGCAACTGGTGATTGTCGTTGCACTGGTGGTTTACGTGATTTATGGATTCGTTGTGGTG TGGGGTGTGTGatgAAATTGCCAGTCAGAGAATTTGATGCAGTTGTGATTGGTGCCGGTGGCGCAGGTATGCGCGCGGCGCTGCAAA TTTCCCAGAGCGGCCAGACCTGTG
groL	GGCTAAATCCACCCGCGGCGAAGTGCTGGCTGTCGGCAATGGCCGTATCCTTGAAAATGGCGAAGTGAAGCCGCTGGATGTGAAAG TTGGCGACATCGTTATTTTCAACGATGGCTACGGTGTGAAATCTGAGAAGATCGACAATGAAGAAGTGTTGATCATGTCCGAAAGCG ACATTCTGGCAATTGTTGAAGCGTAATCCGCGCACGACACTGAACATACGAATTTAAGGAATAAAGATAATGGCAGCTAAAGACGTA AAATTCGGTAACGACGCTCGTGTGAAAATGCTGCGCGGCGTAAACGTACTGGCAGATGCAGTGAAAGTTACCCTCGGTCCAA
ydjA	AATATCACGACAGGAGTTAATCAAATGGATGCACTCGAACTATTGATCAATCGCCGTAGCGCCTCCCGCTTGGCTGAACCCGCGCCAA CGGGTGAACAACTGCAAAACATCCTGCGTGCGGGTA
rspB	TCGATGAAAAGCTGGCGGCGAAATATCCCTATGAACCTGCTTATCTACCAGTCGCACGTCTGGAAGATGGCACGCTGTGGAACTGGT AAGGAGTAAGATAatgAAAAGCATATTAATTGAAAAACCGAATCAACTGGCGATTGTCGAACGTGAAATACCCACCCCGTCAGCGGGT GAAGTACGAGTAAAAGTGAAACTTG
hemG	AATGGATCCTGATTGCCAACATGCTGTTTGGTCGTCTCGAGGTCTTTACATTGCTGGTGCTCTTTACCCCGACTTTCTGGCGTGAATGAT GGAGTAATACGTGAAAACATTAATTCTTTTCTCAACAAGGGACGGACAAACGCGCGAGATTGCCTCCTACCTGGCTTCGGAACTGAAA GAACTGGGGATCCAGGCGGATG
yfgM	GCGCTCTGGTGAGCAAACGGCAGTTGCGCAGGATAGCGTAGCCGCGCATTTGCGCACGTTACTGGGTTAAGGAAGGAGAAGGACAG CGTGGAAATTTACGAGAACGAAAACGACCAGGTAGAAGCGGTTAAACGCTTTTTTGCTGAAAATGGCAAAGCACTGGCTGTTGGGGT GATTTTGGGCGTTG
nhaR	ATACAGCTGGTTACGCGTTCGTTTGCGTCCATCAGTTTGACAGGACGGTTTACCGGGGAGCCATAAACGGCTCCCTTTTCATTGTTATC AGGGAGAGAAAATGAGCATGTCTCATATCAATTACAACCACTTGTATTACTTCTGGCATGTCTATAAAGAAGGTTCCGTGGTTGGCGC AGCGGAGGCGCTTTATTTAACTC
csrA	ATACAGAGAGACCCGACTCTTTTAATCTTTCAAGGAGCAAAGAATGCTGATTCTGACTCGTCGAGTTGGTGAGACCCTCATGATTGGG GATGAGGTCACCGTGACAGTTTTAGGGGTAAAGGGCAACCAGGTACGTATTGGCG
sdiA	AGGGGCGTTGCGGTTTACTATGCAGGATAAGGATTTTTTCAGCTGGCGTCGCACGATGCTGTTGCGTTTTCAGAGGATGGAGACCGC AGAAGAGGTCTACCATGAAATTGAGCTTCAGG
ybeL	AGTTAACGACCCGGGAGATGAAACGATGAACAAGGTTGCTCAATATTACCGTGAACTGGTTGCGTCACTGAGCGAACGCCTGCGCAA TGGCGAACGTGATATCGACGCACTGGTGGAACAGGCGC
ybaL	GTAATTTTGGTTTTCCCGGCCAAAAATGGACAATAACCACCAGGAAAAGGAGACGGAATGCATCACGCCACCCCGCTTATCACCACCA TTGTTGGCGGCCTTGTGCTCGCCTTTATCCTCGGCATGCTGGCCAATAAACTACGTATTTCTCCTCTGG
fdoH	TGCCAATACTTTAACGCCATTCGTTGGTGATGCGAACACGCAGACGCCGGAGTTTAAGTCCTTCCTTGTGAATGTGGAAAAGGTGTAA CGGAGACGACATATGGCTTATCAATCTCAAGATATCATTCGTCGTTCCGCGACTAACGGTCTGACCCCCGCGCCTCAGGCGCGGGACT TCCAGGAAGAAGTGGCGAAACTCA
elaB	TGGTATCCCACACATTGGGATGGCACGCGAGGTAATTCAGGCGTAATCAACAACCCTTGTCTATAGTTAGTGACAGGTTTTACGCAAA

	TGGAGAACGAGAATGTCTAATCAGTTTGGTGATACACGTATCGATGACGACCTGACGCTGCTTAGTGAAACACTGGAAGAGGTGCTC CGCTCCTCTGGCGATCCCGCCGATC
crp	GATGCTACAGTAATACATTGATGTACTGCATGTATGCAAAGGACGTCACATTACCGTGCAGTACAGTTGATAGCCCCTTCCCAGGTAG CGGGAAGCATATTTCGGCAATCCAGAGACAGCGGCGTTATCTGGCTCTGGAGAAAGCTTATAACAGAGGATAACCGCGCATGGTGCT TGGCAAACCGCAAACAGACCCGACTCTCGAATGGTTCTTGTCTCATTGCCACATTCATAAGTACCCATCCAAGAGCACGCTTATTCACC AGG
cysJ	TCTCTTCTGTTTTATGGGCGCTGACAGGGCGCAGAAACAGCTTTGCTTACTGGAACATAACGACGCATGACGACACAGGTCCCACCTT CCGCGTTGCTTCCGTTGAACCCGGAGCAACTGGCACGCCTTCAGGCGGCCACGACCGATTTAACTCCCACCCAGCTTG
cstA	AATGTAACATCTCTATGGACACGCACACGGATAACAACTATGAACAAATCAGGGAAATACCTCGTCTGGACAGTGCTCTCTGTAATGG GAGCATTTGCTCTGGGATACATTGCTTTAAATCGTGGGGAACAGATCAACG
fabB	GAAACTTACTCTATGTGCGACTTACAGAGGTATTGAATGAAACGTGCAGTGATTACTGGCCTGGGCATTGTTTCCAGCATCGGTAATA ACCAGCAGGAAGTCCTGGCATCTCTGCGTGAAGGACGTTCAGGGATCA
dkgA	CACAACACCTCACCGGAGCCTGCTCCGGTGAGTTCATATAAAGGAGGAACGTATGGCTAATCCAACCGTTATTAAGCTACAGGATGGC AATGTCATGCCCCAGCTGGGACTGGGCGTCTGGCAAGCAAGTAATGAGGAAGTAATCACCGCCA
ppc	GACGACGAAAAGCAAAGCCCGAGCATATTCGCGCCAATGCGACGTGAAGGATACAGGGCTATCAAACGATAAGATGGGGTGTCTGG GGTAATATGAACGAACAATATTCCGCATTGCGTAGTAATGTCAGTATGCTCGGCAAAGTGCTGGGAGAAACCATCAAGGATGCGTTG GGAGAACACATTCTTGAAC
pspA	AGCAGGACAATCCTGAACGCAGAAATCAAGAGGACAACATTATGGGTATTTTTTCTCGCTTTGCCGACATCGTGAATGCCAACATCAA CGCTCTGTTAGAGAAAGCGGAAGATCCACAGAAACTGGTTCGTCTGATGATCC
yhiI	GCCGTTTCCTTCTTATCCCGGATGGAAGAATCCGCTGAACGTCGTTCGGCGGCAGTTGTAACACGCGTCGTTTGCTTACAGGGATTAT GGCGGGTCGATTatgGATAAGAGTAAGCGCCATCTGGCGTGGTGGGTTGTCGGGTTACTGGCGGTGGCGGCTATCGTGGCGTGGTGG CTGTTGCGCCCGGCAGGTGTGCCGG
ahpC	GTTGTTGCATTTGTAAGGGCAACACCTCAGCCTGCAGGCAGGCACTGAAGATACCAAAGGGTAGTTCAGATTACACGGTCACCTGGA AAGGGGGCCATTTTACTTTTTATCGCCGCTGGCGGTGCAAAGTTCACAAAGTTGTCTTACGAAGGTTGTAAGGTAAAACTTATCGATTT GATAATGGAAACGCATTAGCCGAATCGGCAAAAATTGGTTACCTTACATCTCATCGAAAACACGGAGGAAGTATAGATGTCCTTGATT AACACCAAAATTAAACCTTTTAAAAACCAGGCATTCAAAAACGGCGAATTCATCGAAATCACCGAAAAAGATACCGAAGGCCGCTGG A
uxaA	CGGCGGACATCAACCTGCTGGGCGAGATGGTGAAAAATATTTGCTTTAACAATGCGCGTGACTACTTCGCCATTGAACTGAACTAAGG TCTGGGGTTGATatgCAATACATCAAGATCCATGCGCTGGATAACGTCGCGGTCGCTTTAGCAGATTTGGCTGAAGGCACAGAAGTCA GTGTCGATAACCAGACTGTTACGC
sucB	CCTCCGCCTCTCCGGCGGTAGGGTATATGTCCGTTCACCAGAAACAGCAACAAGATCTGGTTAATGACGCGCTGAACGTCGAATAAAT AAAGGATACACAatgAGTAGCGTAGATATTCTGGTCCCTGACCTGCCTGAATCCGTAGCCGATGCCACCGTCGCAACCTGGCATAAAA AACCCGGCGACGCAGTCGTACGTG
ppk	ACCCCCGTAATTAAAGCGCCAGCTCTGCCGCTGGCGTTTTTCAATTCACCTGTAAATCGCAAGCTCCAGCAGTTTTTTTCCCCCTTTTCT GGCATAGTTGGACATCTGCCAATATTGCTCGCCATAATATCCAGGCAGTGTCCCGTGAATAAAACGGAGTAAAAGTGGTAATGGGTC AGGAAAAGCTATACATCGAAAAAGAGCTCAGTTGGTTATCGTTCAATGAACGCGTGCTTCAGGAAGCGGCGGACAAATCTAACCCGC TGATTG
sucC	TCTTATCAGGCCTACGGGTGACCGACAATGCCCGGAAGCGATACGAAATATTCGGTCTACGGTTTAAAAGATAACGATTACTGAAGG ATGGACAGAACACATGAACTTACATGAATATCAGGCAAAACAACTTTTTGCCCGCTATGGCTTACCAGCACCGGTGGGTTATGCCTGT ACTACTCCGCGCGAAGCAGAAGAAG
yebE	TAAAAATGGTTATTGAGGAGCAAATATGGCTAACTGGTTAAATCAACTGCAATCCCTGCTTGGGCAAAGCAGTTCTTCTACCTCCTCGT CTGCGGATCAGGGATTGGTCAAACTGTTAGTGCCAG
truC	CCTATTATGAAATGGCGCTGGCGACGGACCATCCGCAACGCGCGCTTATTCTGGCTGAGCTGGAAAAACTGGATGCTCTTTTTGCGGA TGATGCGAGCTAATGCTGGAAATACTCTATCAGGATGAATGGCTGGTTGCGGTAAATAAACCCTCCGGCTGGCTGGTTCACCGCAGCT GGCTGGATCGCGACGAGAAAGTAG
yaeP	GTTTATCTTCGTAGGCTTAGACTTGCATCCACGGTTAAGTCAGAGTGCTGACAGGAGGGCATGTGGAAAAATATTGTGAGTTAATACG CAAGCGGTACGCGGAAATCGCCAGCGGAGACTTAGGATACGTTCCGGACGCGCTGGGCTGCGTTTTGAAAGTGC
glcB	TGCCGTGGGCGTTTCTGGTTTAACCGGAGCACAGGATGCGCAGGTCGCGAAAGCGGCAGCAGCGGTGTTGGCGAAATAAGCGAAAA CGAGGAGATAAACAATGAGTCAAACCATAACCCAGAGCCGTTTACGCATTGACGCCAATTTTAAACGTTTTGTGGATGAAGAAGTTTT ACCGGGAACAGGGCTGGACGCTGCGG
dsrB	AAAGCCACAGGAGGAAAACGATGAAGGTGAATGATCGGGTAACAGTCAAAACGGATGGCGGTCCGCGTCGTCCTGGCGTGGTACTG GCAGTTGAGGAGTTTAGTGAAGGCACAATGTACC
cysK	CCGCATATTCTCTGAGCGGGTATGCTACCTGTTGTATCCCAATTTCATACAGTTAAGGACAGGCCATGAGTAAGATTTTTGAAGATAAC TCGCTGACTATCGGTCACACGCCGCTGGTTCGCCTGAATCGCATCGGTAACGGACGCATTCTGGCGAAGGTGGAAT
lsrF	ATGGTAATGGCATCGTACTGTTACCGGAGCGCGTGATATTCAACAAAGAGAATATCGGCAAATACGATTTCTGATGTGCATTACTTAA CCGGAGTAAGTTATGGCAGATTTAGACGATATTAAAGATGGTAAAGATTTTCGTACCGATCAACCGCAAAAAAATATCCCTTTTACCCT GAAAGGTTGCGGTGCGCTGGATT
ycaC	ATGGTCACAGCAAACTCATATTTCCCGTCTATGCTTTCAGAGCAATGCATCGCGTCTGATCGATTTCAGACGTACCACTGACTCTCGAA AGGAGAAGAGGATGACCAAACCGTATGTTCGTCTTGATAAAAATGATGCTGCCGTTTTGCTTGTTGATCACCAGGCTGGTTTACTTTCC CTTGTACGGGATATCGAACCCG
iscS	CAGAATCAGGCCGGAGTGCTAAATACTCCGTAAACACGGTCGTACATCCAGCCGGTAGCCTGATTCCTTGCATTGAGTGATGTACGGA

	GTTTATAGAGCAATGAAATTACCGATTTATCTCGACTACTCCGCAACCACGCCGGTGGACCCGCGTGTTGCCGAGAAAATGATGCAGT TTATGACGATGGACGGAACCTTTG
yecC	AGAACCAAAATGAGTGCCATTGAAGTTAAGAACCTGGTGAAAAAATTCCACGGTCAGACGGTGCTGCACGGTATCGACCTTGAGGTA AAGCCTGGCGAAGTGGTGGCAA
katG	ACACTGTAGAGGGGAGCACATTGATGAGCACGTCAGACGATATCCATAACACCACAGCCACTGGCAAATGCCCGTTCCATCAGGGCG GTCACGACCAGAGTGCGGGGGCGGGCACAACCACTC
bfr	GAGTGGAAGCGAAGGAGTCAAAAAATGAAAGGTGATACTAAAGTTATAAATTATCTCAACAAACTGTTGGGAAATGAGCTTGTCGCA ATCAATCAGTACTTTCTCCATGCCCGAATGTTTAAAA
csiD	ATGGCTACGAAATGAGCATCGCCATGTCACCCTACATCTCATAAGAGGATCGCTTCTGATGAATGCACTGACCGCCGTACAAAATAAC GCTGTCGATTCAGGCCAGGACTATAGCGGATTCACCCTCACCCCGTCGGCGCAATCCCCGCGTCTGCTGG
rodZ	AACGGTTAACTTAACGGATGTTTCGCGGTGTGGGTGAGCATGATGGTTCACCGGCATCTCAATTCTCATTTAAACGTACCTGCAGCGA ATGAATACTGAAGCCACGCACGACCAAAATGAAGCACTTACTACCGGCGCTCGCCTGCGTAATGCTCGCGAACAACTAGGACTTAGTC AGCAGGCCGTTG
pepT	ACAAAAAGTGAGGGTGACTACATGGATAAACTACTTGAGCGATTTTTGAACTACGTGTCTCTGGATACCCAATCAAAAGCAGGGGTG AGACAGGTTCCCAGCACGGAAGGCCAATGGAAGT
groS	ACCAGCCGGGAAACCACGTAAGCTCCGGCGTCACCCATAACAGATACGGACTTTCTCAAAGGAGAGTTATCAATGAATATTCGTCCAT TGCATGATCGCGTGATCGTCAAGCGTAAAGAAGTTGAAACTAAATCTGCTGGCGGCATCGTTCTGACCGGCTCTGCAGCGGCTA
iscR	GCTATGCAATACCCCCACTTTTACAATAAAAAACCCCGGGCAGGGGCGAGTTTGAGGTGAAGTAAGACatgAGACTGACATCTAAAGG GCGCTATGCCGTGACCGCAATGCTTGACGTTGCGCTCAACTCTGAAGCGGGCCCGGTACCGTTGGCTGATATTTCCGAAC
dnaK	AACCGCAGTGAGTGAGTCTGCAAAAAAATGAAATTGGGCAGTTGAAACCAGACGTTTCGCCCCTATTACAGACTCACAACCACATGAT GACCGAATATATAGTGGAGACGTTTAGATGGGTAAAATAATTGGTATCGACCTGGGTACTACCAACTCTTGTGTAGCGATTATGGATG GCACCACTCCTCGCGTGCTGGAGAACGCCGAAGGCGATC
flu	ACCTGCCGGTATCCACGTTTGTGGGTACCGGCTTTTTTATTCACCCTCAATCTAAGGAAAAGCTGatgAAACGACATCTGAATACCTGCT ACAGGCTGGTATGGAATCACATGACGGGCGCTTTCGTGGTTGCCTCCGAACTGGCCCGCGCACGGGGTAAACGTG
sucA	AGGTTCCTTCGCGAGCCACTACGTAGACAAGAGCTCGCAAGTGAACCCCGGCACGCACATCACTGTGCGTGGTAGTATCCACGGCGA AGTAAGCATAAAAAAGATGCTTAAGGGATCACGATGCAGAACAGCGCTTTGAAAGCCTGGTTGGACTCTTCTTACCTCTCTGGCGCAA ACCAGAGCTGGATAGAACAGCTCTATGAAGACTTCTTAACCGATC
aroG	TATTGCATTCACTAAGATAAGTATGGCAACACTGGAACAGACATGAATTATCAGAACGACGATTTACGCATCAAAGAAATCAAAGAGT TACTTCCTCCTGTCGCATTGCTGGAAAAATTCCCCGCTACTGAAAATGCCGCGA
fabI	ATCTGTTCGTACTGTTTACTAAAACGACGAATCGCCTGATTTTCAGGCACAACAAGCATCAACAATAAGGATTAAAGCTATGGGTTTTC TTTCCGGTAAGCGCATTCTGGTAACCGGTGTTGCCAGCAAACTATCCATCGCCTACGGTATCGCTCAGGCGATGCACCGCGAAGGAGC TG
carB	GCAGTACCGTAAAACCGCTAAGTAATCAGGAGTAAAAGAGCCATGCCAAAACGTACAGATATAAAAAGTATCCTGATTCTGGGTGCG GGCCCGATTGTTATCGGTCAGGCGTGTGAGTTTGACTACTCTGGCGCGCAAGCGT
yceD	TTCGTTTTGCAGGTTGATGTTTGTTATCAGCACTGAACGAAAATAAAGCAGTAACCCGCAATGTGTGCGAATTATTGGCAAAAGGCAA CCACAGGCTGCCTTTTTCTTTGACTCTATGACGTTACAAAGTTAATATGCGCGCCCTATGCAAAAGGTAAAATTACCCCTGACTCTCGAT CCGGTTCGTACGGCTCAAAAACGCCTTGATTACCAGGGTATCTATACCCCTGATCAGGTTGAGCGCG
pntB	GTTGATGTCGGTCACCAACGCGATTTCAGGGATTATTGTTGTCGGAGCACTGTTGCAGATTGGCCAGGGCGGCTGGGTTAGCTTCCTT AGTTTTATCGCGGTGCTTATAGCCAGCATTAATATTTTCGGTGGCTTCACCGTGACTCAGCGCATGCTGAAAATGTTCCGCAAAAATTA AGGGGTAACATATGTCTGGAGGATTAGTTACAGCTGCATACATTGTTGCCGCGATCCTGTTTATCTTCAGTCTGGCCGGTCTTTCGAAA CATGAAACGTCTCGCCAGGGTA
proB	AAATGTGGTAATTTATTAAATCTGTAATAAAAGCGTAAACAACTGCCGCTAGGCTTGCTGATCCCGCGCAACAAAACGCCATGCTTTGC TCGCAGATGGTTGGCAACCGACGACAGTCCTGCTAAAACGTTCGTTTGATATCATTTTTCCTAAAATTGAATGGCAGAGAATCATGAG TGACAGCCAGACGCTGGTGGTAAAACTCGGCACCAGTGTGCTAACAGGCGGATCGCGCCGTCTGAACCGTGCCCATATCGTTGAACT TGTTCGCC
fecA	TTCTCGTTCGACTCATAGCTGAACACAACAAAAATGATGATGGGGAAGGTATGACGCCGTTACGCGTTTTTCGTAAAACAACACCTTT GGTTAACACCATTCGCCTGAGCCTGCTGCCGCTGGCCGGTCTCTCGTTTTCCGCTTTTGCTG
hchA	AGCTCAGTCGCAAATATAGTGACTACCCTAACTAAGCAACAATAAGGAATACACTatgACTGTTCAAACAAGTAAAAATCCGCAGGTCG ATATTGCTGAAGATAATGCATTCTTCCCTTCAGAATATTCGCTTAGCCAATATACCAGTCCTGTCT
yjbD	AAGAGATAAATAGTTAAGAGAAGGCAAAATGAATATTATCAGTTCTGCTCGCAAACGAATTCCCGTACACTGAACTCTCTGGCTCACT GAGGGCACCATCATGGCACTCCCCCGTATTACCCAAAAAGAGATGACTGAACGCGAGCAGCGTGAACTTAAAACACTGCTGGATCGG GCGCGGATTGCGCATGGTCGCGTAT
eno	TTTAACTAGTGACTTGAGGAAAACCTAATGTCCAAAATCGTAAAAATCATCGGTCGTGAAATCATCGACTCCCGTGGTAACCCGACTGT TGAAGCCGAAGTACATCTGGAGGGTGGTTTCGTCGGTA
talA	CTCAGAAGTGTGAATTAACGCACTCATCTAACACTTTACTTTTCAAGGAGTATTTCCTatgAACGAGTTAGACGGCATCAAACAGTTCAC CACTGTCGTGGCAGACAGCGGCGATATTGAGTCCATTCGCCATTATCATCCCCAGGATGCCACCACCA
mreB	GGATTTTCTTTTCCGCCCCAGCTTTCAGGATTATCCCTTAGTATGTTGAAAAAATTTCGTGGCATGTTTTCCAATGACTTGTCCATTGACC TGGGTACTGCGAATACCCTCATTTATGTAAAAGGACAAGGCATCGTATTGA
uxaC	TCTAACTCACGAAAATATCTTCGGACTCTGGAAATTGGTGTGATAACTTTGTCAGCATCGCACCATAAGCAAGCTAGCTCACTCGTTGA GAGGAAGACGAAAatgACTCCGTTTATGACTGAAGATTTCCTGTTAGATACCGAATTTGCCCGCCGTCTGTATCACGACTACGCAAAAG ACCAGCCGATTTTCGATTACCATT

pflB	ATCCACTTAAGAAGGTAGGTGTTACATGTCCGAGCTTAATGAAAAGTTAGCCACAGCCTGGGAAGGTTTTACCAAAGGTGACTGGCA GAATGAAGTAAACGTCCGTGACTTCATTCAGAAAAACT
ydcS	ATAATGCGCGGTAGCTCACAACCTGAATAAATTTTCTCAGGGGCGAAGGTGTGCCTGCAAGCCGCCGTCTATGGTTAAACAAGGAGA TATTTTTACGGCACGGCGGCTGAACAATTAATTACGACAGGAGTAAGACCTTatgAGCAAGACATTTGCCCGCAGCAGCCTGTGTGCG CTCAGCATGACAATAATGACCGCTCACGCCGCCGAACCGCCTACCAATTTAGATAAACCGGAAG
cdd	CATTACATGATTATGAGGCAACGCCATGCATCCACGTTTTCAAACCGCTTTTGCCCAACTTGCGGATAACTTGCAATCTGCACTGGAAC CTATTCTGGCAGACAAGTACTTCCCCGCTTTGTTGA
yeaH	GGTACACATGTTACCGGGCCTACAACGACAGCGAACCGTGGGCCTGAGAAGCGGCAACACAGGCGTAGCATACAGTTGGCAAATGT AGTACGGGGGGCATATGACCTGGTTTATTGACCGGCGTCTGAACGGCAAAAACAAAAGCATGGTGAATCGCCAGCGTTTTTTACGCC GTTATAAAGCGCAAATTAAACAGTCGA
sdhB	AGAGTCGGAATCCATGACGCGCCGAAGCGTCAACATGGAACCGAAACTGCGCCCGGCATTCCCGCCGAAGATTCGTACTTACTAATG CGGAGACAGGAAAatgAGACTCGAGTTTTCAATTTATCGCTATAACCCGGATGTTGATGATGCTCCGCGTATGCAGGATTACACCCTGG AAGCGGATGAAGGTCGCGACATGA
wrbA	AGTGGTAGCGAATCGCTACGGAATAGAGATAACACGAGGAGTGGTTAGAAATGGCTAAAGTTCTGGTGCTTTATTATTCCATGTACG GACATATTGAAACGATGGCACGCGCAGTCGCTGAGGGTGCAAGCAAAGTGGATGGCGCTGAAG
pdxB	GGTTAACTCTCGTCTCATACAGGTAACACAAACGTGAAAATCCTTGTTGATGAAAATATGCCTTATGCCCGCGACTTATTTAGCCGTTT GGGTGAGGTGACCGCGGTTCCCGGGCGTCCAATCCCCGTCGCTC
metC	ACTGCCGTACCTTTGCTTTCTTTTCCTTGCGTTTACGCAGTAAAAAAGTCACCAGCACGCCATTTGCGAAAATTTTCTGCTTTATGCCAAT TCTTCAGGATGCGCCCGCGAATATTCATGCTAGTTTAGACATCCAGACGTATAAAAACAGGAATCCCGACATGGCGGACAAAAAGCTT GATACTCAACTGGTGAATGCAGGACGCAGCAAAAAATACACTCTCGGCGCGGTAAATAGCGTGATTCAGCGCGCTTCTTCGC
lpxC	GAATGTATAGTACACTTCGGTTGGATAGGTAATTTGGCGAGATAATACGATGATCAAACAAAGGACACTTAAACGTATCGTTCAGGCG ACGGGTGTCGGTTTACATACCGGCAAGAAAGTCACCCTGACGTTACGCCCTGCGCCGGCCA
proS	ACTTTTTTTTGTCCCAGGCTCGCCTTGAGCCTGTTCTACCTTCCAACTGGAACCGTAACAACATGCGTACTAGCCAATACCTGCTCTCCA CTCTCAAGGAGACACCTGCCGACGCCGAGGTGATCAGCCATCAGCTGATGCTGCGCGCCGGGATGATCCGCA
glnS	AGATTATCAATTTTAAAAAACTAACAGTTGTCAGCCTGTCCCGCTTATAAGATCATACGCCGTTATACGTTGTTTACGCTTTGAGGAATC CACGATGAGTGAGGCAGAAGCCCGCCCGACTAACTTTATCCGTCAGATCATCGATGAAGATCTGGCCAGTGGTAAGCACACCACAGT ACACACCCGTTTCCCGC
katE	TTTTTTGACCAAAACAGCGGCCCTTTCAGTAATAAATTAAGGAGACGAGTTCAatgTCGCAACATAACGAAAAGAACCCACATCAGCAC CAGTCACCACTACACGATTCCAGCGAAGCGAAACCGGGGATGGACTCACTGGCACCTGAGGACG
ompR	GCTTTTTTAAGAATACACGCTTACAAATTGTTGCGAACCTTTGGGAGTACAAACAATGCAAGAGAACTACAAGATTCTGGTGGTCGAT GACGACATGCGCCTGCGTGCGCTGCTGGAACGTTATCTCACCGAACAAGGCTTCCAGGTTCGAAGCG
guaA	AAGAGTCCCCGAACTACCGTCTGGGCTCCTGATTCTCTTCGCCCGACTTCATGTCGGGCGATTTATATTATCTGTTTCACTTGCCTCGGA ATAAGCGTCAATGACGGAAAACATTCATAAGCATCGCATCCTCATTCTGGACTTCGGTTCTCAGTACACTCAACTGGTTGCGCGCCGCG TGCGTGAGCTGGGTGTTTACT
kdsA	GTAATCCGCAAATTGCTGGATACACTCAAAGCCTCGTTGATGGAAGAAAATCAGATGGAGCTGGCGTTACGCACCAGCGAAGCTTTAT TACAATTCAACCCTGAAGATCCCTATGAAATTCGCGATCGCGGGTTGATTTATGCGCAACTGGATTGCGAACACGTTGCGTTGAACGA TTTAAGTTATTTCGTTGAACAGTGTCCGGAAGACCCGATCAGCGAAATGATCCGTGCGCAAATAAATAACATCGCGCATAAACATATT GTGCTGCATTAATTAATCGACATTTTACTCAAGATTAAGGCGATCCTATGAAACAAAAAGTGGTTAGCATTGGCGACATCAACGTAGC AAATGACCTGCCGTTCGTACTGTTTGGCGGTATGAACGTGTTGGAATCTCGCGATCTGG
aroD	GTTCACATTATGGACTGGCAAAGATTTCCCTCTGGAATATGTTAAACAGGTCATGGGGTTCGGTGCCTGACAGGCTGACCGCGTGCAG AAAGGGTAAAAAatgAAAACCGTAACTGTAAAAGATCTCGTCATTGGTACGGGCGCACCTAAAATCATCGTCTCGCTGATGGCGAAAG ATATCGCCAGCGTGAAATCCGAAG
yqjD	CGCGAACGGAAACTGGCAGAAGCGCAGGAAGAGCTGAAAAAGCTGGAAGCGCGCGACTACTAACTCACAATAGTCACTACTTACTC ACCTGGAGAAAACTATGTCGAAAGAACACACTACGGAACATCTGCGTGCTGAGTTGAAATCCCTTTCCGATACGCTGGAAGAGGTGC TTAGCTCATCTGGCGAGAAGTCGAAAG
pta	AAGACGCGAGCCGCCTGACTGCCTGATTTCACACCGCCAGCTCAGCTGGCGGTGCTGTTTTGTAACCCGCCAAATCGGCGGTAACGAA AGAGGATAAACCGTGTCCCGTATTATTATGCTGATCCCTACCGGAACCAGCGTCGGTCTGACCAGCGTCAGCCTTGGCGTGATCCGTG CAATGGAACGCAAAGGCGTTCGTC
uxuB	GCGCTTTCTTTAGCCGTTAATATCCACCGGCATGGCTGCGCGCCGTGCCGGTTCCTTCTTCCTTGCCGTCACTCTTTGAAGACGGATTCT GGAGTTTACGATGACTACTATTGTTGACAGCAATCTGCCGGTTGCCCGCCCGTCATGGGATCATTCTCGTCTGGAATCACGCATTGTGC ATCTCGGTTGCGGGGCGTTTC
purM	GAATTTTATTTTTCTACCGCAAGTAACGCGTGGGGACCCAAGCAGTGACCGATAAAACCTCTCTTAGCTACAAAGATGCCGGTGTTGA TATTGACGCGGGTAATGCTCTGGTTGGAAGAATCAAAGGCGTAGTGAAGAAAACGC
fhuA	CGTACCGCTTGCGAACCCGCCAGCGTTTCGAATATTATCTTATCTTTATAATAATCATTCTCGTTTACGTTATCATTCACTTTACATCAGA GATATACCAatgGCGCGTTCCAAAACTGCTCAGCCAAAACACTCACTGCGTAAAATCGCAGTTGTAGTAGCCACAGCGGTTAGCGGCA TGTCTGTTTATGCACAGGCAG
cspE	AATCAATAGCTAAAATAAGTAACATCAAAAATAACGCGACTTTTATCACTTTTTAGTAAAGTTACACTGGACAAAGCGTACCACAATTG GTGTACTGGTAACCGACACAGCATTTGTGTCTATTTTTCATGTAAAGGTAATTTTGATGTCTAAGATTAAAGGTAACGTTAAGTGGTTT AATGAGTCCAAAGGATTCGGTTTCATTACTCCGGAAGACGGCAGCAAAGACGTGTTCGTACACTTCT
iaaA	AGTGGCGAACCACCCTTAATGGACGAATACTATGGGCAAAGCAGTCATTGCAATTCATGGTGGCGCAGGTGCAATTAGCCGCGCGCA GATGAGTCTGCAACAGGAATTACGCTACATCGAGGCGTTGTCTG

glmS	GTCACATGGGATGAGGAGATAACATAATCTCCCTCCCACAAGCAGTAACTATAAAAATAACCCCACTCTCTACAAGGCTCGGGGCGCC CGAAAAAACGGGCATACAGGTTGACCGACAACGATATAAATCGGAATCAAAAACTATGTGTGGAATTGTTGGCGCGATCGCGCAACG TGATGTAGCAGAAATCCTTCTTGAAGGTTTACGTCGTCTGGAATACCGCGGATATGACTCTGCCGGTC
mtlD	CCGCATCCGGCGCTGCCCCTCTCCTCACGGAGAGGGTTTGGGTGAGGGAAAAGCCTCACCCCAGCCCTCTCGGGTAAAAACATTGAT GAAGGTTAATACTATGAAAGCATTACATTTTGGCGCAGGTAATATCGGTCGTGGCTTTATCGGTAAACTGCTGGCAGACGCGGGTATC CAACTGACGTTTGCCGATGTCAATC
acs	TTGCGTGATCTGTCGCCCAAATACTAAACAAAACTGCCAATACCCCTACATTTAACGCTTATGCCACATATTATTAACATCCTACAAGGA GAACAAAAGCatgAGCCAAATTCACAAACACACCATTCCTGCCAACATCGCAGACCGTTGCCTGATAAACCCTCAGCAGTACGAGGCG ATGTATCAACAATCTATTAACG
manX	GCGAAACGCAGGGGTTTTTGGTTGTAGCCCTTATCTGAATCGATTCGATTGTGGACGACGATTCAAAAATACATCTGGCACGTTGAGG TGTTAACGATAATAAAGGAGGTAGCAAGTGACCATTGCTATTGTTATAGGCACACATGGTTGGGCTGCAGAGCAGTTGCTTAAAACG GCAGAAATGCTGTTAGGCGAGCAGGAAAACGTCGGCTGGA
suhB	GATTATTCACGCATCTTATCATAAAACGAAGACAGATGCCGATCTCGCTGCTATACTCTGCGCCGTTTTCCCGTTCTTTAACATCCAGTG AGAGAGACCGATGCATCCGATGCTGAACATCGCCGTGCGCGCAGCGCGCAAGGCGGGTAATTTAATTGCCAAAAACTATGAAACCCC GGACGCTGTAGAAGCGAGCCAGA
ackA	TAGGTACTTCCATGTCGAGTAAGTTAGTACTGGTTCTGAACTGCGGTAGTTCTTCACTGAAATTTGCCATCATCGATGCAGTAAATGGT GAAGAGTACCTTTCTGGTTTAG
adeP	TATAAGTAAATCCAATAAAAAAGTCTCTCACAAACGCGCCTATTGTCGCATTTTGGTATACGATAGCGACGGATTTTCCCTCCTTGTTCG GAAATAGATAATGAGTCATCAACATACCACCCAGACTTCTGGTCAGGGGATGCTGGAACGCGTGTTTAAACTGCGCGAACATGGCAC GACGGCACGGACCGAAGTGATCG
adeQ	TCCATTAATGTGCAATTGAAATGTGATAATTATCACTGAACCGTTGTTGTCATTTGTTTTACAAAAGCAAGGGATTAAATCTCCTCCAAT GGACAAAAAAATGAATAATGACAATACCGATTACGTGAGTAATGAATCAGGGACGCTTTCGCGATTATTTAAACTACCTCAGCATGGG ACCACCGTCCGCACAGAATTGA
amyA	ACGGCTACGCTTCTAATGTTCCCCTTGAATGGAGTCGAAGAATGCGTAATCCCACGCTGTTACAATGTTTTCACTGGTATTACCCGGAA GGCGGTAAGCTCTGGCCTGAACTGGCCGAGCGCGCCGACGGTTTTAATGATA
arcA	GTTTTTGACACTGTCGGGTCCTGAGGGAAAGTACCCACGACCAAGCTAATGATGTTGTTGACGTTGATGGAAAGTGCATCAAGAACG CAATTACGTACTTTAGTCATGTTACGCCGATCATGTTAATTTGCAGCATGCATCAGGCAGGTCAGGGACTTTTGTACTTCCTGTTTCGAT TTAGTTGGCAATTTAGGTAGCAAACATGCAGACCCCGCACATTCTTATCGTTGAAGACGAGTTGGTAACACGCAACACGTTGAAAAGT ATTTTCGAAGCGGAAGGCTATGATGTTTTCGAAGCGA
argS	CCCTAATTTCTTTAACTGGTGCGGGCAATTTTTGCTCGCTTCATCAATGTAAGGTATTCCGGTGAATATTCAGGCTCTTCTCTCAGAAAA AGTCCGTCAGGCCATGATTGCGGCAGGCGCGCCTGCGGATTGCGAACCGCAGGTTCGTCAGTCAGCAAAAG
aroA	ATGGTTGAGTTCGAACGCCGTCACGGTTAATGCCGAAATTTTGCTTAATCCCCACAGCCAGCCTGTGGGGTTTTTATTTCTGTTGTAGA GAGTTGAGTTCATGGAATCCCTGACGTTACAACCCATCGCTCGTGTCGATGGCACTATTAATCTGCCCGGTTCCAAGAGCGTTTCTAAC CGCGCTTTATTGCTGGCGGCAT
asd	TTTCACTTGCGACTTTGGCTGCTTTTTGTATGGTGAAAGATGTGCCAAGAGGAGACCGGCACATTTATACAGCACACATCTTTGCAGGA AAAAAACGCTTATGAAAAATGTTGGTTTTATCGGCTGGCGCGGTATGGTCGGCTCCGTTCTCATGCAACGCATGGTTGAAGAGCGCG ACTTCGACGCCATTCGCCCTGTCT
cbpA	AGTTACCTTACAGGGGTTCCTTCAATTTGTGTTGATTTACGCGAGATAACGCTATGGAATTAAAGGATTATTACGCCATCATGGGCGTG AAACCGACGGACGATCTCAAGACAATCAAGACCGCCTATCGTCGACTTGCCCGCAAATACCATC
clpS	AATTTGAAGCAGTTAACGCTATTGACAGGAATGTGACAGATGTCGCTGATGCCAACGATAGATGATAGTTATCTATCATGTGGAGTAG ATTGGTCAGGCAAATAAGCTCTTGTCAGCGGCAGGGCGTTCTGCCGATAACCGTAACCGAAGATGATAACTGACAATGGGTAAAACG AACGACTGGCTGGACTTTGATCAACTGGCGGAAGAAAAAGTTCGCGACGCGCTAAAACCGCCATCTATGTATAAAGTGATATTAGTCA
cyoA	ACACTTTAAACGCCACCAGATCCCGTGGAATTGAGGTCGTTAAATGAGACTCAGGAAATACAATAAAAGTTTGGGATGGTTGTCATTA TTTGCAGGCACTGTATTGCTCAGTGGCTGTAATTCTGCGCTGTTAGATCCCAAAG
cyoB	GACCCAGCCAGAAGGTGAGCACAGCGCACACGAAGGTATGGAAGGCATGGACATGAGCCACGCGGAATCCGCCCATTAAAGGGGTT GAGGAAGAATAAAGATGTTCGGAAAATTATCACTTGATGCAGTCCCGTTCCATGAACCTATCGTCATGGTTACGATCGCTGGCATTAT TTTGGGAGGTCTGGCGCTCGTTGGCC
cysI	TGGAAGTGATTGCCGAATTTGGTGGCATGGACACCGAAGCGGCGGATGAATTTTTAAGTGAGCTGCGCGTAGAGCGCCGTTATCAGC GAGATGTCTACTAATGAGCGAAAAACATCCAGGGCCTTTAGTGGTCGAAGGAAAACTGACAGACGCCGAGCGCATGAAGCATGAAA GCAACTACCTGCGCGGCACCATTGCGG
cysP	GAAAGTCATTAAATTTATAAGGGTGCGCAATGGCCGTTAACTTACTGAAAAAGAACTCACTCGCGCTGGTCGCTTCTCTGCTGCTGGC GGGCCATGTACAGGCAACGGAACTGCTGAACAGTTCTTATG
dacC	TGCGTTATTAATCACCAAACTTATCATACGGCGATATAACGTATTTTTTTTGAATGGATACTCGGGTGGCATTTatgACGCAATACTCCTC TCTCCTTCGTGGTCTTGCAGCGGGTTCTGCATTTTTATTCCTTTTTGCCCCAACGGCATTCGCGGCGGAACAAACCGTTGAAG
dcrB	GTTTCAAAGTGTTAATCAGTATCGTTTTCTTCCTGCCAATGTATGGCGTATTACTCAATATGCTGTTGAAAAGACTGGCAGATAAATCC GAAATCAACGCTTTGCAGGCGAGTTAAAGGTTCGTTATCAGAGTTGTGATAAGATGGATGAATGAGCCGTTATGGCCGTTTATCGAA AGGAAGAAGTCAATGCGCAATCTGGTTAAATATGTCGGAATTGGCCTGCTGGTTATGGGGCTTGCGGCCTGTGATGATAAAGACACT AACGCTACGGCGCAGGGTTCGGTCG
ddlA	TTGTCAGCAGAAGATTGCAACAATCTTTCAGAAGGCCGCGAATCCCCCGTTTTGTCGTGGTAATGTATGCACTTTGTGTTAGACGGGA TAGGTTTTTAAGATGGAAAAACTGCGGGTAGGAATCGTTTTTGGTGGTAAATCAGCGGAACATGAAGTGTCTCTGCAATCGGCAAAA AACATTGTCGATGCCATTGATAAAA

deaD	GATGAGTTATGTAGACTGGCCGCCATTAATTTTGAGGCACACGTACTACatgGCTGAATTCGAAACCACTTTTGCAGATCTGGGCCTGA AGGCTCCTATCCTTGAAGCCCTTAACGATCTGGGTTACGAAAAACCATCTCCAATTCAGG
dgcZ	GCACAAGGAACTGTGAAAAAGGAGTGGCAATGATCAAGAAGACAACGGAAATTGATGCCATCTTGTTAAATCTCAATAAGGCTATCG ATGCCCACTACCAGTGGCTGGTGAGTATGTTTCACAGCGTGG
dppA	ACGAGGGGCATTTTATGGAGGATCCGCACTGTTACACTGATGTTAATTAGTACGGCATCCCCACCTCATAACGTTGACCCGACCGGGC AAAAAACAAAAAAGGTCAGGCAGCGACAACCCACTGCAAAGGGTTAAAACAACAAACATCACAATTGGAGCAGAATAATGCGTATTT CCTTGAAAAAGTCAGGGATGCTGAAGCTTGGTCTCAGCCTGGTGGCTATGACCGTCGCAGCAAGTGTTCAGGCTAAAACTCTGGTTTA TT
fbp	TCCGTAATTTGCTGGCGCGATTCACCTGAGCGCAACATTGAGATTATTTGTTAAGATTGTTGCGGTCGCTTTACTCCATAAACATTGCA GGGAAAGTTTTATGAAAACGTTAGGTGAATTTATTGTCGAAAAGCAGCACGAGTTTTCTCATGCTACCGGTGAGCTCACTGCTTTGCT GTCGGCAATAAAACTGGGCGCCA
feoB	ATTCATATCGAAACCCGTCGTGTGAGCCTGGTATTACGCAAAAAAGATCTGGCCTTATTAGAAGTGGAAGCGGTTTCCTGTTAATACG GTGATAACAACAATGAAAAAATTAACCATTGGCTTAATTGGTAATCCAAATTCTGGCAAGACAACGTTATTTAACCAGCTCACTGGCTC ACGTCAGCGTGTAGGTAACTGGG
fepA	TCTTTCAGGATCAAAGGTTTTCGCGGTAGCGGGATGCGTCGTGTTGATGACGACCATGCCCGACAGTTGCAATTCGTGGCAAAAATGC AGGAATAAAACAatgAACAAGAAGATTCATTCCCTGGCCTTGTTGGTCAATCTGGGGATTTATGGGGTAGCGCAGGCACAAGAGCCGA CCGATACTCCTGTTTCACATGACG
fes	TATATTTCTGCAATCAATGAAAAATTGCACAGTAAACATGGGGTTATGGTGTGACGGCGTTAAAAGTAGGAAGTGAGAGCTGGTGGC AGTCGAAACATGGCCCGGAATGGCAGCGTCTGAATGACGAAATGTTTGAGGTCACTTTCTGGTGGCGTGATCCCCAAGGTTCTGAAG AATACTCGACGATAAAGCGCGTATGGGTCTACATCACTGGTGTGACCGATCACC
flhC	AAATTCATACCGGCATCATGCTCTCAACACGCTTGCTGAATGATGTTAATCAGCCTGAAGAAGCGCTGCGCAAGAAAAGGGCCTGATC ATGAGTGAAAAAAGCATTGTTCAGGAAGCGCGGGATATTCAGCTGGCAATGGAATTGATCACCCTGGGCGCTCGTTTGCAGATGCTG GAAAGCGAAACAC
flhD	GATTTAGGAAAAATCTTAGATAAGTGTAAAGACCCATTTCTATTTGTAAGGACATATTAAACCAAAAAGGTGGTTCTGCTTATTGCAGC TTATCGCAACTATTCTAATGCTAATTATTTTTTACCGGGGCTTCCCGGCGACATCACGGGGTGCGGTGAAACCGCATAAAAATAAAGTT GGTTATTCTGGGTGGGAATAATGCATACCTCCGAGTTGCTGAAACACATTTATGACATCAACTTGTCATATTTACTACTTGCACAGCGT TTGATTGTTCAGGACAAAGCGTCCGCTATGT
fliY	TGATATATATTAAATAAGAATAAGATGTAGCGGAGTTGTTTTTGTGTTTACAAACAATGGCTCTACACTGCAAACAGACATAACAACAT TCGGGGTGAATATGAAATTAGCACATCTGGGACGTCAGGCATTGATGGGTGTGATGGCCGTGGCGCTGGTTGCGGGCATGAGCGTT AAAAGTTTTGCAGATGAAGGTCTGC
frdA	AAAGGAGCAGTGGAATAGCGTTCGCAGACCGTAACTTTCAGGTACTTACCCTGAAGTACGTGGCTGTGGGATAAAAACAATCTGGAG GAATGTCGTGCAAACCTTTCAAGCCGATCTTGCCATTGTAGGCGCCGGTGGCGCGGGATTACGTGCTGCAATTGCTGCCGCGCAGGC AAATCCGAATGCAAAAATCG
frvB	AAAATTATTCACAAGGAATTTATTCATCAACTGCAACAGGGCGATACCGACCAGGTGCTTGCCTTGTTAAATCAAACCCTCAGCTCATA AGGAAGTGGCGATGGAGTCATCCTTACGTATTGTCGCGATCACCAACTGCCCCGCCGGGATCGCTCACACCTACATGGTGGCGGAAG CCCTGGAACAGAAAGCCCGTTCTC
fumA	TTTAACAGGGCAACGGAACACCCGCCCAGAGCATAACCAAACCAGGCAGTAAGTGAGAGAACAatgTCAAACAAACCCTTTCATTATC AGGCTCCTTTTCCACTCAAAAAAGATGATACTGAGTATTACCTGCTAACCAGCGAACACGTTAGCGTATCTGAAT
gabD	GGCGAGATGAAAAACTCGCTGTGCAAAAGCGGCTATCTGCGGCTGGTGCAAAAGTATTGTCCCCGGCTTTCGTTAAGCGATCTCCAG CCCTGGCCCGCCGGTGTGCGGGCGCAGGCGGTATCGCCGGACGGCAAGCTGATTGACGATTTTCTGTTTGTCACCACCCCGCGCACG ATCCACACCTGCAATGCGCCCTCCCCGGCAGCGACATCAGCAATTCCTATTGGTGCGCATATTGTCAGCAAGGTACAAACGCTGTTGG CAAGCCAGAGTAACCCCGGACGCACGCTGCGAGCGGCACGTAGTGTGGATGCCTTACACGCCGCATTTAATCAATAACCTTTGAAAAC AGGATGTAGCGATGAAACTTAACGACAGTAACTTATTCCGCCAGCAGGCGTTGATTAACGGGGAATGGCTGGACGCCAACAATGGTG AAGCCATCGACGTCACCAATCCGG
galM	AAGCAAAAACAGGTATTAAAGAGACTTTTTACGTTTGTAAACCATCACAAGGAGCAGGACAGTGCTGAACGAAACTCCCGCACTGGC ACCCGATGGTCAGCCGTACCGACTGTTAACTTTGCGTAACAACGCAGGGATGGTAGTCACGCTGATGGACTGGG
gatC	TTTTGGCGATATTCCGTTAGTTCACGGCATGCCTTTTATTTCTGGTATCGGTATCGAAGCATTACAAAATAAAATTCTGACTATCTTACA GGGGTGACCTATGTTTTCAGAAGTCATGCGTTATATTCTCGACCTCGGCCCTACGGTGATGCTGCCGATTGTCATCATTATTTTTTCTAA AATATTAGGCATGAAGGCAG
ghrA	GCCTTTGTAGATCATAACGATAAGTGCGAATAAATTTCGCACAACGCTTTTCGGGAGTCAGTATGGATATCATCTTTTATCACCCAACG TTCGATACCCAATGGTGGATTGAGGCACTGCGCAAAGCTATTCCTCAGGCAAGAGTCAGAGCATGGAAAAGCG
glgB	GTGGCTGAATACATGAGTATTCACAGCCTTACCTGAAGTGAGGACGACGCAGAGAGGATGCACAGAGTGCTGCGCCGTTCAGGTCAA AAAAATGTCACAACCAGAAGTCAAAAATCCAATTGGATGGGGTGACACAATAAAACAGGAAGACAAGCATGTCCGATCGTATCGATA GAGACGTGATTAACGCGCTAATTGCAGGCCATTTTGCGGATCCTTTTTCCGTACTGGGAATGCATAAAACCACCGCGGGAC
glnH	GTCACGCAGGGGATCGTCCCGTGGATAGAAAAAAGGAAATGCTATGAAGTCTGTATTAAAAGTTTCACTGGCTGCACTGACCCTGGC TTTTGCGGTTTCTTCTCATGCCGCGGATAAAAAATTAGTTGTCGCGACGGATACCG
gmk	ATGTAGGCTTTATTTCGCTAATCACATACGAAAGATACTCATGGCTCAAGGCACGCTTTATATTGTTTCTGCCCCCAGTGGCGCGGGTA AATCCAGCCTGATTCAGGCTTTATTAAAAACCCAACCGTTGTATGACACCC
gpt	GCGCAACCTATTTTCCCCTCGAACACTTTTTAAGCCGTAGATAAACAGGCTGGGACACTTCACATGAGCGAAAAATACATCGTCACCTG GGACATGTTGCAGATCCATGCACGTAAACTCGCAAGCCGACTGATGCCTTCTGAACAATGGAAAGGCATTATTG
hcaT	CCTGACGGGAGGGACTCATGGTTTTGCAATCCACGCGCTGGTTGGCGCTCGGCTATTTCACATACTTTTTTAGTTACGGCATTTTTCTA

	CCTTTCTGGAGCGTCTGGCTTAAAGGGA
hdeA	GTCGAAATTGATTCGTGACGGCTCTTTCACTTTATAGTTGAGGATATTACGATGAAAAAAGTATTAGGCGTTATTCTTGGTGGTCTGCT TCTTCTGCCAGTTGTGAGCAATGCAGCGGATGCGCAAAAAGCAGCTGATAACAAAAAACCGG
hdeB	ACATCGTACTTCCTTGCATATTGAACAGGCCGGAATATCTTCTTTAAAAGCAGCTATTCCTCCTGTTCATATATAATCTCTATATTGAATG GGTTACAAAatgAATATTTCATCTCTCCGTAAAGCGTTTATTTTTATGGGCGCTGTAGCGGCTTTGTCACTGGTGAACGCACAATCTGCG TTGGCAGCCAATGAATCCG
hflC	GTGGAGCAAGCAACACGTCGTCCACCAGTCAGGGCGATATTATGGACCAACGCCGCGCCAACGCGCAGCGTAACGACTACCAGCGTC AGGGGGAATAACGATGCGTAAGTCAGTTATCGCGATTATCATCATCGTGCTGGTAGTGCTTTACATGTCTGTCTTTGTCGTCAAAGAA GGTGAGCGCGGTATTACGCTGCGTT
hipB	ATATCCCCTTAAGCGGATAAACTTGCTGTGGACGTATGACATGATGAGCTTTCAGAAGATCTATAGCCCAACGCAATTGGCGAATGCA ATGAAACTGGTTCGCCAGCAAAATGGCTGGACGCAGAGCGAGCTGGCGAAAA
hlfC	GTGGAGCAAGCAACACGTCGTCCACCAGTCAGGGCGATATTATGGACCAACGCCGCGCCAACGCGCAGCGTAACGACTACCAGCGTC AGGGGGAATAACGATGCGTAAGTCAGTTATCGCGATTATCATCATCGTGCTGGTAGTGCTTTACATGTCTGTCTTTGTCGTCAAAGAA GGTGAGCGCGGTATTACGCTGCGTT
hyfR	AATTCATTCATCTGTTGGATGAAATTGTGCAGGAACCGGCCATCTATCTGATGGCCAGAAAAATTGCGTGAGAAGGATTTCTCATTAA TAAGGACTGTTGATGGCTATGTCAGACGAGGCGATGTTTGCCCCGCCACAAGGAATAACAATTGAAGCGGTAAACGGAATGCTCGCG GAGCGGTTAGCACAGAAACACGGCA
isnH-7	GTACATTATGCCTGTTCCGAGATGAAGCGAAATCTTTACATGATTATTCCGTAAGAAAATTTCCTGTGCTTTCTGATTTTATTGTGTCAT TTATGTTAGGGATTAAGGAAGGTGCGAACAAGTCCCTGATATGAGATCATGTTTGTCATCTGGAGCCATAGAACAGGGTTCATCATGA GTCATCAACTTACCTTCGCCGACAGTGAATTCAGCAGTAAGCGCCGTCAGACCAGAAAAG
ldcC	AATTCGCAAAAGTTCTGAAAAAGGGTCACTTCGGTGGCCCTTTTTTATCGCCACGGTTTGAGCAGGCTATGATTAAGGAAGGATTTTC CAGGAGGAACACATGAACATCATTGCCATTATGGGACCGCATGGCGTCTTTTATAAAGATGAGCCCATCAAAGAACTGGAGTCGGCG CTGGTGGCGCAAGGCTTTCAGATTA
ldtA	GATAGCTTTCGCGACATAGGAAAGGGACATGATGCGTCGTGTAAATATTCTTTGCTCATTTGCTCTGCTTTTTGCCAGCCATACTAGCC TGGCGGTAACTTATCCATTACCTCCAGAGGGTAGCCGTTTAG
lon	ATGTTAATAGATGGCGTGAAGCACAGTCGTGTCATCTGATTACCTGGCGGAAATTAAACTAAGAGAGAGCTCTATGAATCCTGAGCGT TCTGAACGCATTGAAATCCCCGTATTGCCGCTGCGCGATGTGGTGGTTTATCCGCACATGGTCATCCCCTTATTTGTCGGGCGGG
lrhA	AGTAAGTGATAATATATGATAAGTGCAAATCGTCCGATAATTAACCTCGACCTCGATCTGCTGAGAACATTTGTTGCTGTTGCCGATCT GAACACTTTTGCTGCCGCAGCTGCCG
maeA	GTAAAGCAAAGACGATAAAAGCCCCCCAGGGATGGATATTCAAAAAAGAGTGAGTGACATGGAACCAAAAACAAAAAAACAGCGTT CGCTTTATATCCCTTACGCTGGCCCTGTACTGCTGGAATTTCCGTTGTTGAATAAAGGCAGTGCCTTCAGCA
mdtA	ATTCCGCGAAACGTTTCAGGAAGAGAAACTCTTAACGATGAAAGGTAGTTATAAATCCCGTTGGGTAATCGTAATCGTGGTGGTTATC GCCGCCATCGCCGCATTCTGGTTCTGGCAAGGCCGCAATGACTCCCGGA
mdtE	CTATGCCGCTGGTCTGTAAATCCCTCATATCTCTCCTCGCGCGCAATTTAAAGAACCGTTATTTCTCAAGAATTTTCAGGGACTAAAatg AACAGAAGAAGAAAGCTGTTAATACCGTTGTTATTCTGCGGCGCGATGCTCACCGCCTGCGATGACAAATCGGCGGAAAACGCCGCC GCCATGACGC
purl	GTTTCCCCCCCTTGGGTACACCGAAAGCTTAGAAGACGAGAGACTTATGATGGAAATTCTGCGTGGTTCGCCTGCACTGTCGGCATTC CGAATCAACAAACTGCTGGCACGTTTTCAGGCTGCCAGGCTCCCGGTTCACAATATTT
rbbA	AGCATCTGGAATATGTCAAAACCGGTTTGCCGGGCGTAGCGTGGGTGCGGGTGAATGAAGAACTTCCGTGGCCTGACGACCTCGTGG TGAGGTTGCCGCAATGACGCATCTGGAACTGGTTCCCGTCCCGCCTGTCGCGCAACTGGCGGGCGTGAGCCAGCATTATGGAAAAAC CGTTGCGCTGAACAATATCACTCTCG
ribB	CTGTCGGGCATGGACCCGCTCACGTTATTTTGGCTATATGCCGCCACTCCTAAGACTGCCCTGATTCTGGTAACCATAATTTTAGTGAG GTTTTTTTACCATGAATCAGACGCTACTTTCCTCTTTTGGTACGCCTTTCGAACGTGTTGAAAATGCACTGGCTGCGCTGCGTGAAGGA CGCGGTGTAATGGTGCTTGATG
rpoE	CTTGCTCAAATTGCAGCTAATGGAGTGGCGTTTCGATAGCGCGTGGAAATTTGGTTTGGGGAGACTTTACCTCGGATGAGCGAGCAG TTAACGGACCAGGTCCTGGTTGAACGGGTCCAGAAGGGAGATCAGAAAGCCTTTAACTTACTGGTAGTGCGCTATCAGCATAAAGTG G
rpsR	GTTCAGGGGTTCATTTCATGCCACAAGGCAAAGAACGGACTGAGCAAAATGGTTTTGCATGCCGAGCAGATTGAATTGATAGATTCT GGAGACTAGCCATATGGCACGTTATTTCCGTCGTCGCAAGTTCTGCCGTTTCACCGCGGAAGGCGTTCAAGAGATCGACTATAAAGAT ATCGCTACGCTGAAAAACTACATCA
sapA	AACCACACGGCCCTCATTACGACCCCTGAACACAAACGGTAATGGCAAATGTCATTAAATATTAGTAAGATAGCCTCTTTACATAAAAT CCCTTAATATTATGCGCCAGGTATTATCGTCTCTTTTGGTGATTGCTGGACTTGTGAGTGGTCAGGCAATCGCCGCGCCTGAATCTCCC CCGCATGCTGATATCCGCGACA
slp	AATATTTGTTGATAAGGATAGTAACATGAACATGACAAAAGGTGCACTCATCCTCAGCCTTTCATTTTTGCTTGCCGCATGTAGTTCAA TTCCGCAAAATATCAAAGGCAATAACCAACCTGATA
sucB	CCTCCGCCTCTCCGGCGGTAGGGTATATGTCCGTTCACCAGAAACAGCAACAAGATCTGGTTAATGACGCGCTGAACGTCGAATAAAT AAAGGATACACAatgAGTAGCGTAGATATTCTGGTCCCTGACCTGCCTGAATCCGTAGCCGATGCCACCGTCGCAACCTGGCATAAAA AACCCGGCGACGCAGTCGTACGTG
tam	GTCGCGAAACTGGAATCTTTAATGACCGGGCCGCGTAAAAAACGTCTGTTCAATGGTTTGATGCCGTGAGGCGAATTTATCAATTTTA TCTACAATTGGGGTAACGCGCTGACGGGAGTAAAAAAATGTCTGACTGGAACCCCTCTTTATATCTACACTTTTCCGCTGAACGATCGC GTCCGGCGGTGGAGCTGCTTGCCAGAGTGCCGCTGGAAAATGTCGAAT

tgt	GCAATTAATGAGCGCGTCGGGGAGTAATTCCGCGGCGCTGGTTTAAAACGTTGGACTGTTTTTCTGACGTAGTGGAGAAAAAATGAA ATTTGAACTGGACACCACCGACGGTCGCGCACGCCGTGGCCGCCTGGTCTTTGATCGTGGCGTAGTGGAAACGCCTTGTTTTATGCCT GTTGGCA
thiM	CTGCGATTTATCATCGCAACCAAACGACTCGGGGTGCCCTTCTGCGTGAAGGCTGAGAAATACCCGTATCACCTGATCTGGATAATGC CAGCGTAGGGAAGTCACGGACCACCAGGTCATTGCTTCTTCACGTTATGGCAGGAGCAAACTATGCAAGTCGACCTGCTGGGTTCAG CGCAATCTGCGCACGCGTTACACCTTTTTCACCAACATTCCCCTCTTGTGCACTGCATGACCAATGATGTGGTGC
topA	GTATCGGATTTTATCAGGTACAGTGTGACGCTTTCGTCAATCTGGCAATAGATTTGCTTGACATTCGACCAAAATTCCGTCGTGCTATA GCGCCTGTAGGCCAAGACCTGTTAACTCAGTCACCTGAATTTTCGTGAACAGAGTCACGACAAGGGGTTGATATCCGCAGAGAGCGA GTCCATATCGGTAACTCGTTGCCAGTGGAAGGTTTATCAACGTGCGACGCATTCCTGGAAGAATCAAATTAGGTAAGGTGAATATGG GTAAAGCTCTTGTCATCGTTGAGTCCCCGGCAAAAGCCAAAACGATCAACAAGTATCTGGGTAGTGACTACGTGGTGAAATCCAGCGT CGGTCACA
tyrR	ATTGTTCTTTTTTCAGGTGAAGGTTCCCATGCGTCTGGAAGTCTTTTGTGAAGACCGACTCGGTCTGACCCGCGAATTACTCGATCTAC TCGTGCTAAGAGGCATTGATTTACGCGGTATTGAGATTG
mlaA	AGCAAATCAGGGCGTCTGGACATCAGTTGACGTGCTGTTACAATCGCCCACACCTAAACAGGCGGATACGGTATCGTTCCGTCATGGA TGGCAAACTGCATAAGCCATAAAAAAACAGGGAGACATTTATGAAGCTTCGCCTGTCGGCGCTTGCTCTGGGAACTACGCTTCTGGTG GGGTGTGCGAGTTCCGGTACAGATCAGCAAGGGCGTTCTGACCCGTTAGAAG
nhaA	GGTCACTCGTGAGCGCTTACAGCCGTCAAAAACGCATCTCACCGCTGATGGCGCAAATTCTTCAATAGCTCGTAAAAAACGAATTATTC CTACACTATAATCTGATTTTAACGATGATTCGTGCGGGGTAAAATAGTAAAAACGATCTATTCACCTGAAAGAGAAATAAAAAGTGAA ACATCTGCATCGATTCTTTAGCAGTGATGCCTCGGGAGGCATTATTCTTATCATTGCCGCTATCCTGGCGATGATTATGGCCAACAGCG GCGCAA
nlpA	AATCATTAATAAAAGGATAAAAAAATGAAACTGACAACACATCATCTACGGACAGGGGCCGCATTATTGCTGGCCGGAATTCTGCTG GCAGGTTGCGACCAGAGTAGCAGCGATGCAAAACACA
ntcA/glnA	TACGGCGACACGGCCAAAATAATTGCAGATTTCGTTACCACGACGACCATGACCAATCCAGGAGAGTTAAAGTATGTCCGCTGAACAC GTACTGACGATGCTGAACGAGCACGAAGTGAAGTTTGTTGATTTGCGCTTCACCGATACTAAAGGTAAAGAACAGCACGTCACTA
nuoC	GCCTGTCGGCAAAGGGATTTTTCTTCGCTTATTCCTAAATCTATTTCGCGAAGCTTACTGCGCCGACAGTCACCACGGACCATTTGCAA TGGTGAACAATATGACCGACTTAACCGCGCAAGAACCCGCCTGGCAGACCCGCGATCATCTTGATGATCCGGTGATTGGCGAACTGC GCAACCGTTTTGGGCCGGATGCCT
nuoG	CATTTGATTAATGGGATTCAGCCGAACCTGCTGAAAGAGCGCTGGTAACCGAATTTCGATTAACGCTCAGTCTCTGACTGAGAAAACT GGAAGCATGCTAATGGCTACAATTCATGTAGACGGCAAAGAATACGAGGTCAACGGAGCGGACAACCTGCTGGAAGCTTGTCTGTCT CTGGGCCTTGATATTCCTTACTTTT
oppA	CTGACAGCAGAAAGTCTCCGAGCCTGTGCAGGGTCCCAATCCGGGATTACACATGCTGGTTAATACCAGTAATTATAATGAGGGAGT CCAAAAAACAATGACCAACATCACCAAGAGAAGTTTAGTAGCAGCTGGCGTTCTGGCTGCGCTAATGGCAGGGAATGTCGCGCTGGC AGCTGATGTACCCGCAGGCGTCA
pntA	CGTACATGAGCAGCTTGTGTGGCTCCTGACACAGGCAAACCATCATCAATAAAACCGATGGAAGGGAATATCATGCGAATTGGCATA CCAAGAGAACGGTTAACCAATGAAACCCGTGTTGCAGCAACGCCAAAAACAGTGGAACAGCTGCTGAAACTGGGTTTTACCGTCG
potD	GTAACACAGGGGACGTTAAATGAAAAAATGGTCACGCCACCTGCTCGCGGCGGGTGCTCTGGCACTGGGCATGAGCGCCGCTCACG CCGATGACAACAACACGCTGTATTTCTACAACT
pqqL	TGGCGACCTACGCCAGTGGCCGTACTTTCTGGCTTGATGTCAGCATGAAATTTTAATAGTTCGGATTTTACTTAACCGGGTAAATTCGC CCGGTTTTCGCATGGAGATAATTATGAGAAACCTCTGTTTCTTACTGACGTTAGTGGCAACTCTGTTGCTCCCCGGGCGGCTGATTGCC GCCGCCTTACCGCAGGATGAAA
ptsP	GAAAGAGTTCGCGAGTGTGGTGATGTCACTGCAGGAAAATACGCCAAAACCACAAAACGCATCTGCTTATCGACGTAAAAGAGGTTA AGTCACGCCAATTATGCTCACTCGCCTGCGCGAAATAGTCGAAAAGGTAGCCAGCGCACCACGCCTGAATGAGGCGTTAAATATTCTG GTTACCGACATCTGTCTTGCGATGG
purE	CTCTAAAGCCGAGAGTTGTGCACCACAGGAGTTTTAAGACGCATGTCTTCCCGCAATAATCCGGCGCGTGTCGCCATCGTGATGGGGT CCAAAAGCGACTGGGCTACCATGCAGTTCGCCGCCGAAATCTTCGAAATCCTGA
purK	TTCTTGCGACTCATGATAAAGAACTGCACCAGCGTCTGAATGACTGGCGCAAAGCCCAGACCGACGAAGTGCTGGAAAACCCGGACC CGCGAGGTGCGGCATGAAACAGGTTTGCGTCCTCGGTAACGGGCAGTTAGGCCGTATGCTGCGTCAGGCAGGCGAACCGTTAGGCA TTGCTGTCTGGCCAGTCGGGCTGGACG
yadM	CTTGCGCATTATCAACAAAGCCCAGAAGCAGAAAAAAATAGCACAATCTATTATTCTGAATGGAATTCGGTGTCAGATTCAGTAACTAC AAGGGACGTCAAATGATAAAAACAACGCCACATAAAATAGTGATACTGATGGGAATATTATTATCACCCTCAGTATTTGCAACGGATA TTAATGTAGAGTTTACAGCCACTG
Yaji	AGTGTCCACGGCGAAACAGAATGGGCAATGCATACGGCGTCCTGACCAGGTGGTTAACGGAAATCTATTTTACCCTGAACTGACGAT ACAACAAAGGCGCAACGCTTTTGAGATAAAGGATCGATGCTTTCGCTGCTTTTACCCTCTACCATTAGGCTAATTGCGACGAGAAAGG GACGACCTATGAATACTAACGTTTTTCGACTGCTCCTGCTGGGAAGCCTGTTCAGCCTTAGCGCCTGTGTGCAGCAAAGTGAAGTGCG ACAGATGAAACACAGCGTCA
ybiT	TTTTCAGGATACGCCTGTGTTAGTTTCCAGTAACGTCACCATGCAGTTCGGCAGTAAGCCGTTGTTTGAAAACATTTCCGTCAAATTTG GCGGCGGCAACCGTTACGGCCTGATTG
ybjP	GGATATCATTATTATCGTAGAACGCTTTCAGAGCGATCGCTTATAAGGAAATCATTATGCGCTACAGCAAATTGACAATGCTTATCCCC TGCGCACTGCTCCTCAGTGCCTGCACCACAGTCACTCCAGCTTATAAAGATAACGGCACACGCAGTG
ycaK	AGTCAGACATATTCATCCGCATTAGTGAAAATCCTTCATTAACTCATCCTGAATCATCACGTTTACAGGATTATGCCGCCTGCGTAAAGT GCGCTCCAGAACTTAACGTGGAGGTAAAATTATGCAGTCTGAACGTATTTATTTGGTATGGGCCCATCCTCGTCATGATTCATTGACCG

	CACATATTGCTGATGCGATCCATCAGCGGGCAATGGAGCGGA
ycdT	AAAGGGATCTACAACCTACAGATTGGTGTAGCTTTATGGAAAAAGACTATTTGAGAATTAGTAGTACTGTATTAGTGAGCTTATTGTTT GGGCTTGCTTTGGTTTTGGTGAATAGTTGGTTTAATCAGCCAGGCG
ydcJ	ATTAAAGCAATTATGTTACAGCAAAATGGATAATATTGATGTTTTCGCGGCGAGATCACAGTTTGTAAATTCTTCCCGCAAGAGTGAAT GCGGTTACCTACACTCCAGATTACTGACCACTGGAGGCAGACACTatgGCGAACAGCATCACGGCGGATGAGATTCGGGAACAGTTTT CGCAGGCAATGTCAGCCATGTACCAGCAAGAAGTTCCGCAATATGGCACGCTGCTGG
ydgA	AGGCTTATAACACCTTCAGGCGGCCAGTCCGCCTGATTTCATTTTATGGATAATCATTATGAATAAATCGCTGGTAGCGGTAGGCGTCA TTGTTGCGCTAGGCGTAGTCTGGACAGGCGGCGCATGGTATACAGGCAAGAAGATTGAAACCCATCTCG
yeaG	ACGTTGTTCTCATCGTCGATAAAATGGCATGAGAGTTGCTGTGTTTTAGCAAGAGACGTCGTTCAGTTTACCTCTTCCGGGAGCCTCTA CTATTCATATGAACGGCTCTTAACCTGTGCTAAAAAACGAAAGGACGGCATACCATGAATATATTCGATCACTATCGCCAGCGATATG AAGCTGCCAAGGACGAAGAGTTCACACTGCAGGAGTTTCTTACCACTTGTCGGCAAGATCGCAGTG
yeaY	TATGCCCTGAGAAAAAAAGGAGTCGCACCATGGCGGTTCAAAAGAATGTTATCAAAGGCATACTGGCAGGTACGTTTGCGCTAATGC TGAGCGGTTGTGTCACTGTGCCGGACGCCATTAAAGGCAGCA
yebF	ATCCGAAGAGGAAGATGCCTCACACGCGGCCTGAAGCAACTTGCGCTCCTGACAATTTGTACCATGCTTTCTCCTGAATCATTGTGCTA TGGAGAAAAACATGAAAAAAAGAGGGGCGTTTTTAGGGCTGTTGTTGGTTTCTGCCTGCGCATCAGTTTTCGCTGCCAATAATGAAAC CAGCAAGTCGGTCACTTTCCCAA
yfhM	CGACGTTGCTAACGGGCCATTGTGGCAGGGATGAAAAATGAAAAAGTTACGCGTAGCCGCCTGCATGCTAATGCTGGCGCTGGCAG GGTGCGACAACAACGATAACGCGCCAACAGCGGTGAAAAAAGATGCGCCTT
yhbO	TTGCGCTCCACGGCTCATGCACCAGGCTTAATAGACCATGAGGTAATTATCCCCGATTGTGGGAAATTCGCCTCATCCAATGCAACAAC GCGGAGGAAGCATGAGTAAGAAAATTGCCGTTTTAATCACTGATGAATTTGAGGATTCAGAATTTACTTCACCCGCAGACGAGTTCCG TAAAGCCGGACACGAAGTGATTA
yhcB	GAGATGTTCATGACCTGGGAATATGCGCTAATTGGGTTAGTCGTCGGCATCATTATTGGTGCTGTGGCCATGCGTTTTGGTAATCGTA AACTACGCCAGCAACAGGCGT
yhjG	CATTTTCCGCCTGGCTGGACTATCTTTAGGACTGGCACAGGAAGAATATGAGGAAGCGAACGATGAGCAAGGCAGGCAAAATAACCG CTGCGATTTCAGGGGCTTTCTTGTTGTTGATTGTCGTGGCGATCATTTTGATTGCAACTT
yiaD	GGCAGTCAGAACGGAATATCGTGAGGATGAGTTGGCCAAACCGAAGACTCCTGAAAGGTTAGATGTATTGAGCAGTTAAATATTACT TACAAATTTCCGAGTATTCCCAGGAATAATCTTCATGTTCACGCGGCATAATCTCCCGCCACGGAACCCGTGGCAAAGAATAAAAAGG TTATTAAGGATTAACAATGAAGAAACGTGTTTATCTTATTGCCGCCGTAGTGAGTGGCGCTCTGGCGGTATCTGGCTGCACAACTAAC CCTTACACCGGCGAACGCGAAGCAGGTA
yihX	CAAAAATGCTCTATATCTTTGATTTAGGTAATGTGATTGTCGATATCGACTTTAACCGTGTGCTGGGAGCCTGGAGCGATTTAACGCGT ATTCCGCTGGCATCGC
yiiS	ATGGTAATCCATAAGATCATTACTTGTTTGTTCTTCCCTTAACGGCGCATTATTCTTAAGTCGTAATCGACAGAGAGGCGAATATACAG AGGTGCCCTATGAAAGATGTCGTAGATAAATGCAGTACTAAAGGATGTGCGATTGATATCGGTACGGTGATTGATAACGACAACTGT ACCAGTAAGTTTTCGCGCTTTT
yliE	ATGTATTAACATGCTGAGTTTATACGAAAAGATAAAGATAAGGCTGATAATTTTATTTTTATTGGCAGCACTGTCATTTATTGGTCTTTT TTTCATCATT
yliF	AAATTAGGATGTGTTCAGGCTCAGGGGTATCTGTACCAGAAACCCATGCCATTCTCCGCCTGGGATAAAAGTGGAAAATTAGTAAAA GAGTAGTTTACGTATGTCCAGAATCAATAAGTTCGTACTTACAGTCAGTCTGCTGATTTTTATCATGATTTCAGCAGTTGCCTGCGGGA TCTACACTCAAATGGTAAAGGAAC
yqjE	CGCGTGCCGATGAGTATGTGCGCGAAAATCCGTGGACGGGCGTGGGCATTGGCGCTGCAATCGGTGTAGTGCTCGGCGTTCTGCTGT CGCGTCGTTAATTATGGCGGACACTCATCACGCACAAGGGCCCGGTAAAAGCGTTCTGGGCATCGGGCAGCGAATTGTTTCTATCATG GTTGAAATGGTAGAGACACGTCTGC
yqjG	ATAAAAAACGAGGAGGAAGCTCCTCGTTTTTGCTATTGGAGGAGAGAAAAATGGGTCAACTGATTGACGGCGTCTGGCATGACACCT GGTACGATACCAAATCTACCGGCGGTAAATTTCAACGTTCAGCTTCCGCATTTCGTAACTGGC
ytfQ	ATGAAAGCGATTACAAACTTGTGATTAACGTTTTATTTACTTTTTTGAAGTGTGATGTAACGCAATCTGTTACATAACGAATTGTCTATA GTTTTTTCGCGAACATCTTTTAACCAATAATAACTACCCCGACGAGGACAACCCTATGTGGAAACGCTTACTTATAGTCTCTGCAGTCTC GGCAGCCATGTCGTCTATGGCGTTGGCCGCTCCATTAACCGTTGGATTTTCGCAGGTCGGATCGG

Table A.2 Primer design for genomic PCRs and cloning technique for tested genes

Gene	Plasmid generated by	Forward Primer	Reverse Primer
pgaA	Restriction Digest SalI/SphI	ACACGTCGACAGGCATTGGGATTTATGCCG	TTTACGCATGCGGCACCTTTTTCTGCTACTTG
nhaR	Restriction Digest SalI/SphI	ACACAGGTCGACATACAGCTGGTTACGCGTTC	TTTACGCATGCGAGTTAAATAAAGCGCCTCCG
dgcZ	Genscript		
galM	Restriction Digest SalI/SphI	CAGTAGTCGACAAGCAAAAACAGGTATTAAAGAGACTTTTTACG	AACTGAGCATGCCCCAGTCCATCAGCGTGA
gadA	Genscript		
cbpA	Gibson	TGGATTCCACACAGGTCGACAGTTACCTTACAGGGGTTC	AAGTTCTTCTCCTTTACGCATGCGATGGTATTTGCGGGCAAG
groS	Restriction Digest SalI/SphI	ATAGTAGTCGACACCAGCCGGGAAACC	TCAAGCATGCTAGCCGCTGCAGAGC
groL	Restriction Digest SalI/SphI	CAGTAGTCGACGGCTAAATCCACCCGCG	AACTGAGCATGCTTGGACCGAGGGTAACTTTCAC
aidB	Restriction Digest SalI/SphI	CAGTAGTCGACGTGACTGCCATTGATGGAGGG	AACTGAGCATGCCACGCGTTACCGCTTCGC
gstA	Genscript		
ucpA	Genscript		
yhiI	Restriction Digest SalI/SphI	CAGTAGTCGACGCCGTTTCCTTCTTATCCCG	AACTGAGCATGCCCAGTAACCCGACAACCCAC
rbbA	Genscript		
mdtA	Restriction Digest SalI/SphI	CAGTAGTCGACATTCCGCGAAACGTTTCAGG	AACTGAGCATGCTCCGGGAGTCATTGCG
yfhM	Gibson	TGGATTCCACACAGGTCGACCGACGTTGCTAACGGGCC	AAGTTCTTCTCCTTTACGCATGCAAGGCGCATCTTTTTTCACCG
sdhA	Restriction Digest SalI/SphI	CAGTAGTCGACACGTTAAACCGCTGGCTTTG	AACTGAGCATGCCACAGGTCTGGCCGCTC
cysI	Restriction Digest SalI/SphI	CAGTAGTCGACTGGAAGTGATTGCCGAATTTGG	AACTGAGCATGCCCGCAATGGTGCCGC
cysJ	Restriction Digest SalI/SphI	ATAGTAGTCGACTCTCTTCTGTTTTATGGGCGC	TCAAGCATGCGATTTAACTCCCACCCAGCTTG
yebE	Restriction Digest SalI/SphI	CAGTAGTCGACTAAAAATGGTTATTGAGGAGCAAATATGGC	AACTGAGCATGCCTGGCACTAACAGTTTGACCAATC
ybaL	Genscript		

maeB	Genscript		
fucO	Restriction Digest SalI/SphI	CAGTAGTCGACAGTGCTGAGCGATGAAGAG	AACTGAGCATGCGCGCCTTCTGATAACCACG
mdtE	Restriction Digest SalI/SphI	CAGTAGTCGACCTATGCCGCTGGTCTGTAAATCC	AACTGAGCATGCCGTTTTCCGCCGATTTGTCATC
sdhB	Restriction Digest SalI/SphI	CAGTAGTCGACAGAGTCGGAATCCATGACGC	AACTGAGCATGCTCATGTCGCGACCTTCATCC
tauD	Restriction Digest SalI/SphI	CAGTAGTCGACTGCTGGCGGGGATCG	AACTGAGCATGCGTTCAAACTGATTATCGCTTAACGGG
ydeP	Restriction Digest SalI/SphI	CAGTAGTCGACATGAGAGATGATGCTTAACTTATCGTA	AACTGAGCATGCCCTGGCGTATATCCATCTGC
yaji	Restriction Digest SalI/SphI	CAGTAGTCGACAGTGTCCACGGCGAAACA	AACTGAGCATGCTGACGCTGTGTTTCATCTGTC
pspA	Restriction Digest SalI/SphI	CAGTAGTCGACAGCAGGACAATCCTGAACG	AACTGAGCATGCGGATCATCAGACGAACCAGTTTC
rpoS	Restriction Digest SalI/SphI	CAGTAGTCGACGGAAATCCGTAAACCCGCTG	AACTGAGCATGCCCTGTTCTACTAAGGCCTTTTTCG
sucA	Restriction Digest SalI/SphI	CAGTAGTCGACAGGTTCCTTCGCGAGCCA	AACTGAGCATGCGATCGGTTAAGAAGTCTTCATAGAGC
sucB	Restriction Digest SalI/SphI	CAGTAGTCGACCCTCCGCCTCTCCGG	AACTGAGCATGCCACGTACGACTGCGTCGC
yihX	Gibson	TGGATTCCACACAGGTCGACCAAAAATGCTCTATATCTTTGATTTAGGTAATG	AAGTTCTTCTCCTTTACGCATGCGCGATGCCAGCGGAATAC
sucC	Restriction Digest SalI/SphI	ATAGTAGTCGACTCTTATCAGGCCTACGGGTG	TCAAGCATGCCTTCTTCTGCTTCGCGCG
glsA	Gibson	TGGATTCCACACAGGTCGACGGATGCCTGAAAATCGGC	AAGTTCTTCTCCTTTACGCATGCCCAGAAAGGGAATGTAATCG
ydjA	Restriction Digest SalI/SphI	CAGTAGTCGACAATATCACGACAGGAGTTAATCAAATGG	AACTGAGCATGCTACCCGCACGCAGGAT
iscS	Genscript		
ltaE	Restriction Digest SalI/SphI	CAGTAGTCGACATTTCCCGTCATAATAAGGACATGC	AACTGAGCATGCGGTCGTCTCCGTAAACGTCG
ldtA	Gibson	TGGATTCCACACAGGTCGACGATAGCTTTCGCGACATAG	AAGTTCTTCTCCTTTACGCATGCCTAAACGGCTACCCTCTG
uspG	Genscript		
yebF	Gibson	TGGATTCCACACAGGTCGACATCCGAAGAGGAAGATGC	AAGTTCTTCTCCTTTACGCATGCTTGGGAAAGTGACCGACTTG
amyA	Restriction Digest	ATAGTAGTCGACACGGCTACGCTTCTAATGTTCC	TCAAGCATGCTATCATTAAAACCGTCGGCGCG

	SalI/SphI		
hdeB	Restriction Digest SalI/SphI	CAGTAGTCGACACATCGTACTTCCTTGCATATTGAAC	AACTGAGCATGCCGGATTCATTGGCTGCCAAC
hdeA	Restriction Digest SalI/SphI	CAGTAGTCGACGTCGAAATTGATTCGTGACGGC	AACTGAGCATGCCCGGTTTTTTGTTATCAGCTGC
poxB	Restriction Digest SalI/SphI	CAGTAGTCGACGGCTATTTAACCGTTAGTGCCTC	AACTGAGCATGCCACTAAGACCGTTCAGAGAGTCG
hemG	Gibson	TGGATTCCACACAGGTCGACAATGGATCCTGATTGCCAACATG	AAGTTCTTCTCCTTTACGCATGCCATCCGCCTGGATCCCCA
evgA	Restriction Digest SalI/SphI	CAGTAGTCGACATTATCTTAAAGGAAGCTCAGATTTTCTT	AACTGAGCATGCCTTCAGTCAACTCTGCTAAGATTTC
dnaK	Gibson	TGGATTCCACACAGGTCGACAACCGCAGTGAGTGAGTC	AAGTTCTTCTCCTTTACGCATGCGATCGCCTTCGGCGTTCT
asd	Gibson	TGGATTCCACACAGGTCGACTTTCACTTGCGACTTTGGCTG	AAGTTCTTCTCCTTTACGCATGCAGACAGGGCGAATGGCGT
hemX	Gibson	TGGATTCCACACAGGTCGACCGCCCGGGAACTGGGCTG	AAGTTCTTCTCCTTTACGCATGCTGTTCTTACTCTTTTTTTCTGTTGCGACAGGTTG
gpt	Gibson	TGGATTCCACACAGGTCGACGCGCAACCTATTTTCCCC	AAGTTCTTCTCCTTTACGCATGCCAATAATGCCTTTCCATTGTTC
cstA	Restriction Digest SalI/SphI	ACACGTCGACAATGTAACATCTCTATGGACACGC	AACTGAGCATGCCGTTGATCTGTTCCCCACGAT
gadB	Restriction Digest SalI/SphI	CAGTAGTCGACCGATCCAATCATTTTAAGGAGTTTAAAATGG	AACTGAGCATGCGCGGAAAACGTTTTGATTCTGC
hflC	Gibson	TGGATTCCACACAGGTCGACGTGGAGCAAGCAACACGT	AAGTTCTTCTCCTTTACGCATGCAACGCAGCGTAATACCGC
ldcC	Restriction Digest SalI/SphI	CAGTAGTCGACAATTCGCAAAAGTTCTGAAAAAGGG	AACTGAGCATGCTAATCTGAAAGCCTTGCGCC
pgm	Restriction Digest SalI/SphI	CAGTAGTCGACAGCATGATCGAACACATCATGC	AACTGAGCATGCCTGCTTCTGGTTTCAGTACATAATATTGCG
nuoG	Gibson	TGGATTCCACACAGGTCGACCATTTGATTAATGGGATTCAGC	AAGTTCTTCTCCTTTACGCATGCAAAAGTAAGGAATATCAAGGCC
yhjG	Genscript		
glpR	Genscript		
cyoB	Gibson	TGGATTCCACACAGGTCGACGACCCAGCCAGAAGGTGAG	AAGTTCTTCTCCTTTACGCATGCGGCCAACGAGCGCCAGAC
yqjD	Restriction Digest SalI/SphI	ATAGTAGTCGACCGCGAACGGAAACTGGC	TCAAGCATGCCTTTCGACTTCTCGCCAGATG
yqjE	Gibson	TGGATTCCACACAGGTCGACCGCGTGCCGATGAGTATG	AAGTTCTTCTCCTTTACGCATGCGCAGACGTGTCTCTACCATTTC
acnA	Restriction Digest SalI/SphI	ACACAGGTCGACAACTGTTTGCTGAAGATGATCAGC	TTTACGCATGCCGCCCAGTGATTTAGCAGC
dacC	Restriction	CAGTAGTCGACTGCGTTATTAATCACCAAAC	AACTGAGCATGCCTTCAACGGTTTGTTCCG

	Digest SalI/SphI	TTATCATACGG	CC
frdA	Genscript		
fumA	Restriction Digest SalI/SphI	CAGTAGTCGACTTTAACAGGGCAACGGAACAC	AACTGAGCATGCATTCAGATACGCTAACGTGTTCGC
gatC	Gibson	TGGATTCCACACAGGTCGACTTTTGGCGATATTCCGTTAG	AAGTTCTTCTCCTTTACGCATGCCTGCCTTCATGCCTAATATTTTAG
mscS	Genscript		
pntB	Gibson	TGGATTCCACACAGGTCGACGTTGATGTCGGTCACCAAC	AAGTTCTTCTCCTTTACGCATGCTACCCTGGCGAGACGTTT
dps	Restriction Digest SalI/SphI	CAGTAGTCGACGTTAATTACTGGGACATAACATCAAGAGG	AACTGAGCATGCTCAGCAACTCTACTGTTGCTTTTTTC
ydgA	Restriction Digest SalI/SphI	CAGTAGTCGACAGGCTTATAACACCTTCAGGCG	AACTGAGCATGCCGAGATGGGTTTCAATCTTCTTGC
slp	Restriction Digest SalI/SphI	CAGTAGTCGACAATATTTGTTGATAAGGATAGTAACATGAACATG	AACTGAGCATGCTATCAGGTTGGTTATTGCCTTTGAT
ydhQ	Gibson	TGGATTCCACACAGGTCGACGTCGTGGATAAATCCTCC	AAGTTCTTCTCCTTTACGCATGCTCAACTCGCCTTCAGTAAG
elaB	Gibson	TGGATTCCACACAGGTCGACTGGTATCCCACACATTGGGATG	AAGTTCTTCTCCTTTACGCATGCGATCGGCGGGATCGCCAG
yhbO	Genscript		
bfr	Restriction Digest SalI/SphI	ATAGTAGTCGACGAGTGGAAGCGAAGGAGTC	TCAAGCATGCTTTTAAACATTCGGGCATGGAGAAAG
proP	Gibson	TGGATTCCACACAGGTCGACGTCATTAACTGCCCAATTCAGG	AAGTTCTTCTCCTTTACGCATGCTACCCAGTGATGCTGCGG
fbp	Gibson	TGGATTCCACACAGGTCGACTCCGTAATTTGCTGGCGC	AAGTTCTTCTCCTTTACGCATGCTGGCGCCCAGTTTTATTG
talA	Restriction Digest SalI/SphI	CAGTAGTCGACCTCAGAAGTGTGAATTAACGCACTC	AACTGAGCATGCTGGTGGTGGCATCCTGG
glgB	Gibson	TGGATTCCACACAGGTCGACGTGGCTGAATACATGAGTATTC	AAGTTCTTCTCCTTTACGCATGCGTCCCGCGGTGGTTTTAT
ybjP	Gibson	TGGATTCCACACAGGTCGACGGATATCATTATTATCGTAGAACGCTTTC	AAGTTCTTCTCCTTTACGCATGCCACTGCGTGTGCCGTTAT
ycaC	Gibson	TGGATTCCACACAGGTCGACATGGTCACAGCAAACTCATATTTC	AAGTTCTTCTCCTTTACGCATGCCGGGTTCGATATCCCGTAC
maeA	Restriction Digest SalI/SphI	CAGTAGTCGACTTTGTTTTTATCTGCTTTATACTTGAGGC	AACTGAGCATGCTGCTGAAGGCACTGCC
dcrB	Gibson	TGGATTCCACACAGGTCGACGTTTCAAAGTGTTAATCAGTATCGTTTTCTTCCTGCC	AAGTTCTTCTCCTTTACGCATGCCGACCGAACCCTGCGCCG
yeaY	Gibson		
fdoH	Genscript		
ahr	Restriction Digest SalI/SphI	CAGTAGTCGACAACAACACCAGAGAAGGACC	AACTGAGCATGCGCACTTCAACATCTTGTGGC

frdB	Gibson	TGGATTCCACACAGGTCGACGACGTGAAGATTACTACGC	AAGTTCTTCTCCTTTACGCATGCTTGCGTCATAAGGCACTTC
nuoC	Restriction Digest SalI/SphI	CAGTAGTCGACGCCTGTCGGCAAAGGG	AACTGAGCATGCAGGCATCCGGCCCAAAAC
ddlA	Restriction Digest SalI/SphI	CAGTAGTCGACTTGTCAGCAGAAGATTGCAAC	AACTGAGCATGCTTTTATCAATGGCATCGACAATGTT
hflK	Genscript		
dkgA	Restriction Digest SalI/SphI	CAGTAGTCGACCACAACACCTCACCGGAG	AACTGAGCATGCTGGCGGTGATTACTTCCTC
hchA	Restriction Digest SalI/SphI	CAGTAGTCGACAGCTCAGTCGCAAATATAGTGACT	AACTGAGCATGCAGACAGGACTGGTATATTGGCT
wrbA	Restriction Digest SalI/SphI	CAGTAGTCGACAGTGGTAGCGAATCGCTACG	AACTGAGCATGCCTTCAGCGCCATCCACTTTG
glcB	Gibson	TGGATTCCACACAGGTCGACTGCCGTGGGCGTTTCTGG	AAGTTCTTCTCCTTTACGCATGCCCGCAGCGTCCAGCCCTG
ybeL	Restriction Digest SalI/SphI	ATAGTAGTCGACAGTTAACGACCCGGGAGATG	TCAAGCATGCGCGCCTGTTCCACCAG
mlaA	Gibson	TGGATTCCACACAGGTCGACAGCAAATCAGGGCGTCTG	AAGTTCTTCTCCTTTACGCATGCCTTCTAACGGGTCAGAACG
eno	Restriction Digest SalI/SphI	CAGTAGTCGACTTTAACTAGTGACTTGAGGAAAACC	AACTGAGCATGCTACCGACGAAACCACCC
yhcB	Gibson	TGGATTCCACACAGGTCGACGAGATGTTCATGACCTGGGAATATGC	AAGTTCTTCTCCTTTACGCATGCACGCCTGTTGCTGGCGTA
gabD	Gibson	TGGATTCCACACAGGTCGACGGCGAGATGAAAAACTCGC	AAGTTCTTCTCCTTTACGCATGCCCGGATTGGTGACGTCGA
pck	Genscript		
dsrB	Gibson	TGGATTCCACACAGGTCGACAAAGCCACAGGAGGAAAAC	AAGTTCTTCTCCTTTACGCATGCGGTACATTGTGCCTTCAC
clpB	Restriction Digest SalI/SphI	ATAGTAGTCGACAGTAGCAACTTTGATCCGTTATGGG	TCAAGCATGCAATGAAGTGGTTCGATAAATTGGTTGTC
yiiS	Gibson	TGGATTCCACACAGGTCGACATGGTAATCCATAAGATCATTACTTG	AAGTTCTTCTCCTTTACGCATGCAAAAGCGCGAAAACTTAC
cyoA	Gibson	TGGATTCCACACAGGTCGACACACTTTAAACGCCACCAG	AAGTTCTTCTCCTTTACGCATGCCTTTGGGATCTAACAGCG
oppA	Restriction Digest SalI/SphI	CAGTAGTCGACCTGACAGCAGAAAGTCTCCGAGC	AACTGAGCATGCTGACGCCTGCGGGTACAT
pta	Restriction Digest SalI/SphI	CAGTAGTCGACAAGACGCGAGCCGCCT	AACTGAGCATGCGACGAACGCCTTTGCGTTCC
mtlD	Gibson	TGGATTCCACACAGGTCGACCCGCATCCGGCGCTGCCC	AAGTTCTTCTCCTTTACGCATGCGATTGACATCGGCAAACGTCAGTTGGATACCC
yeaG	Gibson	TGGATTCCACACAGGTCGACACGTTGTTCTCATCGTCG	AAGTTCTTCTCCTTTACGCATGCCACTGCGATCTTGCCGAC

lsrF	Gibson	TGGATTCCACACAGGTCGACATGGTAATGGCATCGTACTGTTAC	AAGTTCTTCTCCTTTACGCATGCAATCCAGCGCACCGCAAC
ytfQ	Gibson	TGGATTCCACACAGGTCGACATGAAAGCGATTACAAACTTGTGATTAAC	AAGTTCTTCTCCTTTACGCATGCCCGATCCGACCTGCGAAA
yeaP	Gibson	TGGATTCCACACAGGTCGACGTTTATCTTCGTAGGCTTAGAC	AAGTTCTTCTCCTTTACGCATGCGCACTTTCAAAACGCAGC
fepA	Restriction Digest SalI/SphI	CAGTAGTCGACTCTTTCAGGATCAAAGGTTTTCGC	AACTGAGCATGCCGTCATGTGAAACAGGAGTATCG
aroG	Restriction Digest SalI/SphI	CAGTAGTCGACTATTGCATTCACTAAGATAAGTATGGCAACAC	AACTGAGCATGCTCGCGGCATTTTCAGTAGCG
clpA	Gibson	TGGATTCCACACAGGTCGACACGTGCAACGCAATTGATG	AAGTTCTTCTCCTTTACGCATGCGCGCCAGTAACAAGTGCT
aroA	Restriction Digest SalI/SphI	CAGTAGTCGACATGGTTGAGTTCGAACGCC	AACTGAGCATGCATGCCGCCAGCAATAAAGC
kdsA	Restriction Digest SalI/SphI	CAGTAGTCGACGTAATCCGCAAATTGCTGGATACAC	AACTGAGCATGCCCAGATCGCGAGATTCCAACA
fliY	Restriction Digest SalI/SphI	CAGTAGTCGACTGATATATATTAAATAAGAATAAGATGTAGCGGAGTTG	AACTGAGCATGCGCAGACCTTCATCTGCAAAACTTTTAAC
ybiT	Restriction Digest SalI/SphI	CAGTAGTCGACTTTTCAGGATACGCCTGTGTTAGTTTC	AACTGAGCATGCCAATCAGGCCGTAACGGTTGC
yceD	Restriction Digest SalI/SphI	CAGTAGTCGACTTCGTTTTGCAGGTTGATGTTTG	AACTGAGCATGCCGCGCTCAACCTGATCAG
purK	Gibson	TGGATTCCACACAGGTCGACTTCTTGCGACTCATGATAAAGAACTGCACC	AAGTTCTTCTCCTTTACGCATGCCGTCCAGCCCGACTGGCC
potD	Genscript		
guaA	Restriction Digest SalI/SphI	CAGTAGTCGACAAGAGTCCCCGAACTACCG	AACTGAGCATGCAGTAAACACCCAGCTCACG
purL	Genscript		
entF	Restriction Digest SalI/SphI	CAGTAGTCGACTTGTGTGTCAGCCGCAGTC	AACTGAGCATGCCAACGTAATGCGCCACGCT
carB	Gibson	TGGATTCCACACAGGTCGACGCAGTACCGTAAAACCGCTAAGTAATCAG	AAGTTCTTCTCCTTTACGCATGCACGCTTGCGCGCCAGAGT
yecC	Gibson	TGGATTCCACACAGGTCGACAGAACCAAAATGAGTGCCATTGAAG	AAGTTCTTCTCCTTTACGCATGCTTGCCACCACTTCGCCAG
pdxB	Gibson	TGGATTCCACACAGGTCGACGGTTAACTCTCGTCTCATACAGG	AAGTTCTTCTCCTTTACGCATGCGAGCGACGGGGATTGGAC
ppk	Gibson	TGGATTCCACACAGGTCGACACCCCCGTAATTAAAGCG	AAGTTCTTCTCCTTTACGCATGCCAATCAGCGGGTTAGATTTG
thiG	Gibson	TGGATTCCACACAGGTCGACGCGATTAATCAGCAAATCGTCCCGC	AAGTTCTTCTCCTTTACGCATGCCGGAAGCGCGGATCGCCT
uxaC	Restriction Digest SalI/SphI	CAGTAGTCGACTCTAACTCACGAAAATATCTTCGGACTC	AACTGAGCATGCAATGGTAATCGAAAATCGGCTGG

ycaK	Genscript		
aroD	Restriction Digest SalI/SphI	CAGTAGTCGACGTTCACATTATGGACTGGCAAAGATTTC	AACTGAGCATGCCTTCGGATTTCACGCTGGCG
hipB	Genscript		
yjbD	Gibson	TGGATTCCACACAGGTCGACAAGAGATAAATAGTTAAGAGAAGGCAAAATG	AAGTTCTTCTCCTTTACGCATGCATACGCGACCATGCGCAA
rnk	Gibson	TGGATTCCACACAGGTCGACGAGTTAAACCCTCGCCGCC	AAGTTCTTCTCCTTTACGCATGCACGCGTCGGCGATTGGCA
purM	Restriction Digest SalI/SphI	CAGTAGTCGACGAATTTTATTTTTCTACCGCAAGTAACG	AACTGAGCATGCGCGTTTTCTTCACTACGCC
uxuA	Genscript		
uxuB	Restriction Digest SalI/SphI	CAGTAGTCGACGCGCTTTCTTTAGCCGTTAATATCC	AACTGAGCATGCGAAACGCCCCGCAACC
purE	Gibson	TGGATTCCACACAGGTCGACCTCTAAAGCCGAGAGTTG	AAGTTCTTCTCCTTTACGCATGCTCAGGATTTCGAAGATTTCG
mreB	Genscript		
proS	Restriction Digest SalI/SphI	CAGTAGTCGACACTTTTTTTTGTCCCAGGCTC	AACTGAGCATGCTGCGGATCATCCCGGC
entC	Gibson	TGGATTCCACACAGGTCGACGAAAATATAAATGATAATCATTATTAAAGCC	AAGTTCTTCTCCTTTACGCATGCCTGACGTCGTAAAACTGC
ribB	Gibson	TGGATTCCACACAGGTCGACCTGTCGGGCATGGACCCG	AAGTTCTTCTCCTTTACGCATGCCATCAAGCACCATTACACCGC
glmS	Gibson	TGGATTCCACACAGGTCGACGTCACATGGGATGAGGAG	AAGTTCTTCTCCTTTACGCATGCGACCGGCAGAGTCATATC
cdd	Gibson	TGGATTCCACACAGGTCGACCATTACATGATTATGAGGCAACG	AAGTTCTTCTCCTTTACGCATGCTCAACAAAGCGGGGAAGTAC
csiD	Restriction Digest SalI/SphI	CAGTAGTCGACATGGCTACGAAATGAGCATCG	AACTGAGCATGCCCAGCAGACGCGGGG
yqjG	Restriction Digest SalI/SphI	CAGTAGTCGACATAAAAAACGAGGAGGAAGCTCC	AACTGAGCATGCGCCAGTTACGAAATGCGGAAG
uxaB	Restriction Digest SalI/SphI	CAGTAGTCGACATTCAAACAGGTTGTATGACTAATCAGAAG	AACTGAGCATGCGCCAGTCAACAAAGGCGC
truC	Restriction Digest SalI/SphI	CAGTAGTCGACCCTATTATGAAATGGCGCTGGC	AACTGAGCATGCCTACTTTCTCGTCGCGATCCAG
moaB	Restriction Digest SalI/SphI	CAGTAGTCGACGCGGGAGAAGAAACAGACCC	AACTGAGCATGCGCAGATAGTGACCGGAGGTATC
flu	Restriction Digest SalI/SphI	CAGTAGTCGACACCTGCCGGTATCCACGT	AACTGAGCATGCCACGTTTACCCCGTGCG
uxaA	Restriction Digest SalI/SphI	CAGTAGTCGACCGGCGGACATCAACCT	AACTGAGCATGCGCGTAACAGTCTGGTTATCGAC

fecA	Gibson	TGGATTCCACACAGGTCGACTTCTCGTTCGACTCATAGC	AAGTTCTTCTCCTTTACGCATGCCAGCAAAAGCGGAAAACG
gmk	Gibson	TGGATTCCACACAGGTCGACATGTAGGCTTTATTTCGCTAATC	AAGTTCTTCTCCTTTACGCATGCGGGTGTCATACAACGGTTG
flhD	Restriction Digest SalI/SphI	ACACGTCGACGATTTAGGAAAAATCTTAGATAAGTGTAAAGACCC	TTTTACGCATGCGAGTCAGTGCCCGCTAACG
ycdT	Genscript		
sdiA	Genscript		
ompC	Genscript		
phoU	Genscript		
gadC	Genscript		
osmE	Genscript		
yebY	Gibson	TGGATTCCACACAGGTCGACGTTTGTGTGGATGACTAAAC	AAGTTCTTCTCCTTTACGCATGCTACCCACTTCAAAACGAC
patA	Genscript		
glgC_short_UTR	Genscript		0
pgaA	Restriction Digest SpeI/SphI	CACTACTAGTAGGCATTGGGATTTATGCCG	TTTACGCATGCGGCACCTTTTTCTGCTACTTG
nhaR	Restriction Digest SpeI/SphI	CACTACTAGTATACAGCTGGTTACGCGTTCG	TTTACGCATGCGAGTTAAATAAAGCGCCTCCG
dgcZ	Genscript		0
gadA	Genscript		0
groS	Restriction Digest SpeI/SphI	CTATACTAGTACCAGCCGGGAAACC	TCAAGCATGCTAGCCGCTGCAGAGC
groL	Restriction Digest SpeI/SphI	CACTACTAGTGGCTAAATCCACCCGCG	AACTGAGCATGCTTGGACCGAGGGTAACTTTCAC
aidB	Restriction Digest SpeI/SphI	CACTACTAGTGTGACTGCCATTGATGGAGG	AACTGAGCATGCCACGCGTTACCGCTTCGC
gstA	Restriction Digest SpeI/SphI	TTAAATTTGCGCACACTAGTAGGGAAACTCTTTTACAAGC	GCTCACACGTCTTCGCATGCTCATTAAATCCACACTGACG
yhiI	Restriction Digest SpeI/SphI	CACTACTAGTGCCGTTTCCTTCTTATCCCG	AACTGAGCATGCCCAGTAACCCGACAACCCAC
sdhA	Restriction Digest SpeI/SphI	CACTACTAGTACGTTAAACCGCTGGCTTTG	AACTGAGCATGCCACAGGTCTGGCCGCTC
yebE	Restriction Digest SpeI/SphI	CACTACTAGTTAAAAATGGTTATTGAGGAGCAAATATGGC	AACTGAGCATGCCTGGCACTAACAGTTTGACCAATC
maeB	Genscript		0
fucO	Restriction	CACTACTAGTAGTGCTGAGCGATGAAGAG	AACTGAGCATGCGCGCCTTCTGATAACCAC

	Digest SpeI/SphI		G
tauD	Restriction Digest SpeI/SphI	CTATACTAGTTGCTGGCGGGGATCG	AACTGAGCATGCGTTCAAACTGATTATCGC TTAACGGG
rpoS	Restriction Digest SpeI/SphI	CACTACTAGTGGAAATCCGTAAACCCGCTG	AACTGAGCATGCCCTGTTCTACTAAGGCCTT TTCG
sucB	Restriction Digest SpeI/SphI	CACTACTAGTCCTCCGCCTCTCCGG	AACTGAGCATGCCACGTACGACTGCGTCGC
sucC	Restriction Digest SpeI/SphI	CTATACTAGTTCTTATCAGGCCTACGGGTG	TCAAGCATGCCTTCTTCTGCTTCGCGCG
glsA	Restriction Digest SpeI/SphI	TTAAATTTGCGCACACTAGTGGATGCCTGAA AATCGGC	AAGTTCTTCTCCTTTACGCATGCCCAGAAAG GGAATGTAATCG
ydjA	Restriction Digest SpeI/SphI	CACTACTAGTAATATCACGACAGGAGTTAAT CAAATGG	AACTGAGCATGCTACCCGCACGCAGGAT
ltaE	Restriction Digest SpeI/SphI	CACTACTAGTATTTCCCGTCATAATAAGGAC ATGC	AACTGAGCATGCGGTCGTCTCCGTAAACGT CG
ldtA	Gibson	TTAAATTTGCGCACACTAGTGATAGCTTTCG CGACATAG	GCTCACACGTCTTCGCATGCCTAAACGGCT ACCCTCTG
uspG	Genscript		0
amyA	Restriction Digest SpeI/SphI	CTATACTAGTACGGCTACGCTTCTAATGTTCC	TCAAGCATGCTATCATTAAAACCGTCGGCG CG
poxB	Restriction Digest SpeI/SphI	CACTACTAGTGGCTATTTAACCGTTAGTGCC TC	AACTGAGCATGCCACTAAGACCGTTCAGAG AGTCG
evgA	Gibson	TTAAATTTGCGCACACTAGTTACTTGTCGAA TTATCTTAAAGGAAG	GCTCACACGTCTTCGCATGCCTTCAGTCAAC TCTGCTAAG
hemX	Restriction Digest SpeI/SphI	CTATACTAGTCGCCCGGGAACTGG	AAGTTCTTCTCCTTTACGCATGCTGTTCTTAC TCTTTTTTTCTGTTGCGACAGGTTG
cstA	Restriction Digest SpeI/SphI	CACTACTAGTAATGTAACATCTCTATGGACA CGCA	AACTGAGCATGCCGTTGATCTGTTCCCCAC GAT
gadB	Gibson	TTAAATTTGCGCACACTAGTCGATCCAATCA TTTTAAGGAG	GCTCACACGTCTTCGCATGCGCGGAAAACG TTTTGATTC
pgm	Restriction Digest SpeI/SphI	CACTACTAGTAGCATGATCGAACACATCATG C	AACTGAGCATGCCTGCTTCTGGTTTCAGTAC ATAATATTGCG
nuoG	Restriction Digest SpeI/SphI	CTATACTAGTCATTTGATTAATGGGATTCAG CCG	AAGTTCTTCTCCTTTACGCATGCAAAAGTAA GGAATATCAAGGCC
yqjE	Restriction Digest SpeI/SphI	CTATACTAGTCGCGTGCCGATGAGTATGTG	AAGTTCTTCTCCTTTACGCATGCGCAGACGT GTCTCTACCATTTC
acnA	Restriction Digest	CACTACTAGTTCACCCTGAAGAGAATCAGGG	TTTACGCATGCCGCCCAGTGATTTAGCAGC

	SpeI/SphI		
dps	Gibson	TTAAATTTGCGCACACTAGTGTTAATTACTGGGACATAACATC	GCTCACACGTCTTCGCATGCTCAGCAACTCTACTGTTG
ydhQ	Restriction Digest SpeI/SphI	CTATACTAGTGTCGTGGATAAATCCTCCGTAG	AAGTTCTTCTCCTTTACGCATGCTCAACTCGCCTTCAGTAAG
elaB	Restriction Digest SpeI/SphI	CTATACTAGTTGGTATCCCACACATTGGGATG	AAGTTCTTCTCCTTTACGCATGCGATCGGCGGGATCGCCAG
proP	Restriction Digest SpeI/SphI	CTATACTAGTGTCATTAACTGCCCAATTCAGGCG	AAGTTCTTCTCCTTTACGCATGCTACCCAGTGATGCTGCGG
fbp	Restriction Digest SpeI/SphI	CTATACTAGTTCCGTAATTTGCTGGCGC	AAGTTCTTCTCCTTTACGCATGCTGGCGCCCAGTTTTATTG
talA	Restriction Digest SpeI/SphI	CACTACTAGTCTCAGAAGTGTGAATTAACGCACTC	AACTGAGCATGCTGGTGGTGGCATCCTGG
fdoH	Restriction Digest SpeI/SphI	CTATACTAGTTGCCAATACTTTAACGCCATTCG	TCAAGCATGCTGAGTTTCGCCACTTCTTCC
ahr	Restriction Digest SpeI/SphI	CACTACTAGTAACAACACCAGAGAAGGACC	AACTGAGCATGCGCACTTCAACATCTTGTGGC
frdB	Restriction Digest SpeI/SphI	CTATACTAGTGACGTGAAGATTACTACGCTGC	AAGTTCTTCTCCTTTACGCATGCTTGCGTCATAAGGCACTTC
nuoC	Restriction Digest SpeI/SphI	CACTACTAGTGCCTGTCGGCAAAGGG	AACTGAGCATGCAGGCATCCGGCCCAAAAC
hflK	Genscript		0
dkgA	Restriction Digest SpeI/SphI	CACTACTAGTCACAACACCTCACCGGAG	AACTGAGCATGCTGGCGGTGATTACTTCCTC
hchA	Restriction Digest SpeI/SphI	CACTACTAGTAGCTCAGTCGCAAATATAGTGACT	AACTGAGCATGCAGACAGGACTGGTATATTGGCT
wrbA	Restriction Digest SpeI/SphI	CACTACTAGTAGTGGTAGCGAATCGCTACG	AACTGAGCATGCCTTCAGCGCCATCCACTTTG
glcB	Restriction Digest SpeI/SphI	CTATACTAGTTGCCGTGGGCGTTTCTG	AAGTTCTTCTCCTTTACGCATGCCCGCAGCGTCCAGCCCTG
ybeL	Restriction Digest SpeI/SphI	CTATACTAGTAGTTAACGACCCGGGAGATG	TCAAGCATGCGCGCCTGTTCCACCAG
eno	Restriction Digest SpeI/SphI	CACTACTAGTTTTAACTAGTGACTTGAGGAAAACCTAATG	AACTGAGCATGCTACCGACGAAACCACCC
pck	Restriction Digest SpeI/SphI	AACTGAGCATGCTACCGACGAAACCACCC	AACTGAGCATGCGATACAGCAGGTCGTAGC

dsrB	Restriction Digest SpeI/SphI	CTATACTAGTAAAGCCACAGGAGGAAAACG	AAGTTCTTCTCCTTTACGCATGCGGTACATTGTGCCTTCAC
clpB	Restriction Digest SpeI/SphI	CTATACTAGTAGTAGCAACTTTGATCCGTTATGGG	TCAAGCATGCAATGAAGTGGTTCGATAAATTGGTTGTC
yiiS	Restriction Digest SpeI/SphI	CTATACTAGTATGGTAATCCATAAGATCATTACTTGTTTGTTC	AAGTTCTTCTCCTTTACGCATGCAAAAGCGCGAAAACTTAC
pta	Restriction Digest SpeI/SphI	CACTACTAGTAAGACGCGAGCCGCCT	AACTGAGCATGCGACGAACGCCTTTGCGTTCC
lsrF	Restriction Digest SpeI/SphI	CTATACTAGTATGGTAATGGCATCGTACTGTTACC	AAGTTCTTCTCCTTTACGCATGCAATCCAGCGCACCGCAAC
ytfQ	Restriction Digest SpeI/SphI	CTATACTAGTATGAAAGCGATTACAAACTTGTGA	AAGTTCTTCTCCTTTACGCATGCCCGATCCGACCTGCGAAA
fepA	Restriction Digest SpeI/SphI	CACTACTAGTTCTTTCAGGATCAAAGGTTTTCGC	AACTGAGCATGCCGTCATGTGAAACAGGAGTATCG
kdsA	Restriction Digest SpeI/SphI	CTATACTAGTGTAATCCGCAAATTGCTGGATACAC	AACTGAGCATGCCCAGATCGCGAGATTCCAACA
ybiT	Restriction Digest SpeI/SphI	CTATACTAGTTTTTCAGGATACGCCTGTGTTAGTTTC	AACTGAGCATGCCAATCAGGCCGTAACGGTTGC
guaA	Restriction Digest SpeI/SphI	CTATACTAGTAAGAGTCCCCGAACTACCG	AACTGAGCATGCAGTAAACACCCAGCTCACG
entF	Restriction Digest SpeI/SphI	CACTACTAGTTTGTGTGTCAGCCGCAG	AACTGAGCATGCCAACGTAATGCGCCACGCT
carB	Restriction Digest SpeI/SphI	CTATACTAGTGCAGTACCGTAAAACCGCTAAG	AAGTTCTTCTCCTTTACGCATGCACGCTTGCGCGCCAGAGT
thiG	Restriction Digest SpeI/SphI	CTATACTAGTGCGATTAATCAGCAAATCGTCCC	AAGTTCTTCTCCTTTACGCATGCCGGAAGCGCGGATCGCCT
uxaC	Restriction Digest SpeI/SphI	CACTACTAGTTCTAACTCACGAAAATATCTTCGGACTC	AACTGAGCATGCAATGGTAATCGAAAATCGGCTGG
aroD	Restriction Digest SpeI/SphI	CACTACTAGTGTTCACATTATGGACTGGCAAAGATTTC	AACTGAGCATGCCTTCGGATTTCACGCTGGCG
yjbD	Restriction Digest SpeI/SphI	CTATACTAGTAAGAGATAAATAGTTAAGAGAAGGCAAAATG	AAGTTCTTCTCCTTTACGCATGCATACGCGACCATGCGCAA
purM	Restriction Digest SpeI/SphI	CACTACTAGTGAATTTTATTTTTCTACCGCAAGTAACG	AACTGAGCATGCGCGTTTTCTTCACTACGCC
proS	Restriction Digest	CACTACTAGTACTTTTTTTTGTCCCAGGCTC	AACTGAGCATGCTGCGGATCATCCCGGC

	SpeI/SphI		
ribB	Restriction Digest SpeI/SphI	CTATACTAGTGCTTATTCTCAGGGCGGGG	AAGTTCTTCTCCTTTACGCATGCCATCAAGCACCATTACACCGC
cdd	Restriction Digest SpeI/SphI	CTATACTAGTCATTACATGATTATGAGGCAACGCC	AAGTTCTTCTCCTTTACGCATGCTCAACAAAGCGGGGAAGTAC
fecA	Restriction Digest SpeI/SphI	CACTACTAGTTTCTCGTTCGACTCATAGCTGAAC	AAGTTCTTCTCCTTTACGCATGCCAGCAAAAGCGGAAAACG
ompC	Genscript		
phoU	Genscript		
gadC	Restriction Digest SpeI/SphI	CTATACTAGTGTGCGGTTCCGACAGGAATAC	TCAAGCATGCTAGGGTATTCATAAACAGCCATTACC
osmE	Genscript		
yebY	Restriction Digest SpeI/SphI	CTATACTAGTGTTTGTGTGGATGACTAAACTCG	AAGTTCTTCTCCTTTACGCATGCTACCCACTTCAAAACGAC
patA	Genscript		
csiD	Restriction Digest SpeI/SphI	cactactagtATGGCTACGAAATGAGCATCG	aactgaGCATGCCCAGCAGACGCGGGG
moaB	Restriction Digest SalI/SphI	CAGTAGTCGACGCGGGAGAAGAAACAGACCC	AACTGAGCATGCGCAGATAGTGACCGGAGGTATC
ybaL	Restriction Digest SpeI/SphI	ctatactagtGTAATTTTGGTTTTCCCGGCC	tcaagcatgcCCAGAGGAGAAATACGTAGTTTATTGG
yhjG	Restriction Digest SpeI/SphI	ctatactagtCATTTTCCGCCTGGCTGG	tcaagcatgcAAGTTGCAATCAAAATGATCGCC
bfr	Restriction Digest SpeI/SphI	ctatactagtGAGTGGAAGCGAAGGAGTC	tcaagcatgcTTTTAAACATTCGGGCATGGAGAAAG
yecC	Restriction Digest SpeI/SphI	ctatactagtAGAACCAAAATGAGTGCCATTGAAG	tcaagcatgcTTGCCACCACTTCGCCAG
hfq	Genscript		
hfq	Restriction Digest SpeI/SphI	cactactagtGTATCGTGCGCAATTTTTTCAG	tTTACGCaTGCaCTCGATTTGCCCTTGCAGC
cysP	Restriction Digest SpeI/SphI	CAGTAGTCGACGAAAGTCATTAAATTTATAAGGGTGCGC	AACTGAGCATGCCATAAGAACTGTTCAGCAGTTCCG
rpsR	Restriction Digest SpeI/SphI	CAGTAGTCGACATTCTGGAGACTAGCCATATGGC	AACTGAGCATGCTGATGTAGTTTTTCAGCGTAGCG
lon	Restriction Digest SpeI/SphI	CAGTAGTCGACATGTTAATAGATGGCGTGAAGCAC	AACTGAGCATGCCCCGCCCGACAAATAAGG

gstB	Restriction Digest SpeI/SphI	CAGTAGTCGACCTGTTCAGCCTGTGGCTG	AACTGAGCATGCGGCCCGCGAGAATTTGC
tgt	Restriction Digest SpeI/SphI	CAGTAGTCGACGCAATTAATGAGCGCGTCGG	AACTGAGCATGCTGCCAACAGGCATAAAACAAGG
glnS	Restriction Digest SpeI/SphI	CAGTAGTCGACAGATTATCAATTTTAAAAAACTAACAGTTGTCAGCC	AACTGAGCATGCGCGGGAAACGGGTGTGTAC
pflB	Restriction Digest SpeI/SphI	CAGTAGTCGACATCCACTTAAGAAGGTAGGTGTTACATG	AACTGAGCATGCAGTTTTTCTGAATGAAGTCACGGACG
lrhA	Restriction Digest SpeI/SphI	CAGTAGTCGACAGTAAGTGATAATATATGATAAGTGCAAATCGTCC	AACTGAGCATGCCGGCAGCTGCGGCA
cmk	Restriction Digest SpeI/SphI	CAGTAGTCGACTTATGTTAACGGTACGCCTGT	AACTGAGCATGCGCAGATGCCATTGCAACG
ompR	Restriction Digest SpeI/SphI	CAGTAGTCGACGCTTTTTTAAGAATACACGCTTACAAATTG	AACTGAGCATGCCGCTTCGAACCTGGAAGC
rodZ	Restriction Digest SpeI/SphI	CAGTAGTCGACAACGGTTAACTTAACGGATGTTTCG	AACTGAGCATGCCAACGGCCTGCTGACTAAG
crp	Restriction Digest SpeI/SphI	CAGTAGTCGACGATGCTACAGTAATACATTGATGTACTGC	AACTGAGCATGCCCTGGTGAATAAGCGTGCTC
cspE	Restriction Digest SpeI/SphI	CAGTAGTCGACAATCAATAGCTAAAATAAGTAACATCAAAAATAACGC	AACTGAGCATGCAGAAGTGTACGAACACGTCTTTG
yeaH	Restriction Digest SpeI/SphI	CAGTAGTCGACGGTACACATGTTACCGGGC	AACTGAGCATGCTCGACTGTTTAATTTGCGCTTTATAACG
feoB	Restriction Digest SpeI/SphI	CAGTAGTCGACATTCATATCGAAACCCGTCGTG	AACTGAGCATGCCCCAGTTACCTACACGCTG
pqqL	Restriction Digest SpeI/SphI	CAGTAGTCGACTGGCGACCTACGCCAG	AACTGAGCATGCTTTCATCCTGCGGTAAGGC
nnr	Restriction Digest SpeI/SphI	CAGTAGTCGACCGCTAACTGATTGAGATCGAGGG	AACTGAGCATGCGCACATCTGCCGCCTC
katG	Restriction Digest SpeI/SphI	CAGTAGTCGACACACTGTAGAGGGGAGCAC	AACTGAGCATGCGAGTGGTTGTGCCCGC
tyrR	Restriction Digest SpeI/SphI	CAGTAGTCGACATTGTTCTTTTTTCAGGTGAAGGTTCC	AACTGAGCATGCCAATCTCAATACCGCGTAAATCAATGC
lpxC	Restriction Digest SpeI/SphI	CAGTAGTCGACGAATGTATAGTACACTTCGGTTGGATAGG	AACTGAGCATGCGTAACGTCAGGGTGACTTTCTTGC
clpS	Restriction Digest	CAGTAGTCGACAATTTGAAGCAGTTAACGCTATTGACAG	AACTGAGCATGCTGACTAATATCACTTTATACATAGATGGCGG

	SpeI/SphI		
iaaA	Restriction Digest SpeI/SphI	CAGTAGTCGACAGTGGCGAACCACCCT	AACTGAGCATGCCAGACAACGCCTCGATGTAG
manX	Restriction Digest SpeI/SphI	CAGTAGTCGACGCGAAACGCAGGGGT	AACTGAGCATGCTCCAGCCGACGTTTTCC
proB	Restriction Digest SpeI/SphI	CAGTAGTCGACAAATGTGGTAATTTATTAAATCTGTAATAAAAGCG	AACTGAGCATGCGGCGAACAAGTTCAACGATATG
pepT	Restriction Digest SpeI/SphI	CAGTAGTCGACACAAAAAGTGAGGGTGACTACATG	AACTGAGCATGCACTTCCATTGGCCTTCCG
topA	Restriction Digest SpeI/SphI	CAGTAGTCGACGTATCGGATTTTATCAGGTACAGTGTG	AACTGAGCATGCTGTGACCGACGCTGGATT
metC	Restriction Digest SpeI/SphI	CAGTAGTCGACACTGCCGTACCTTTGCTTTC	AACTGAGCATGCGCGAAGAAGCGCGCTG
astD	Restriction Digest SpeI/SphI	CAGTAGTCGACCCGAGCGTTTGATTTTAACCGC	AACTGAGCATGCCATCATTGCCTTGCCATAACAC
fes	Restriction Digest SpeI/SphI	CAGTAGTCGACTATATTTCTGCAATCAATGAAAAATTGCACAGTAAAC	AACTGAGCATGCGGTGATCGGTCACACCAGTG
iaaA	Restriction Digest SpeI/SphI	CAGTAGTCGACAGTGGCGAACCACCCT	AACTGAGCATGCCAGACAACGCCTCGATGTAG
clpS	Restriction Digest SpeI/SphI	CAGTAGTCGACAATTTGAAGCAGTTAACGCTATTGACAG	AACTGAGCATGCTGACTAATATCACTTTATACATAGATGGCGG
manX	Restriction Digest SpeI/SphI	CAGTAGTCGACGCGAAACGCAGGGGT	AACTGAGCATGCTCCAGCCGACGTTTTCC
suhB	Restriction Digest SpeI/SphI	CACTGTCGACGATTATTCACGCATCTTATCATAAAACGAAG	AAGTAGCATGCTCTGGCTCGCTTCTACAGC

Table A.3 TriFC plasmid modification

TriFC sequence	DNA sequence
speI-RBS-sphI-mstrawberry-aatII	ACTAGTAGGAGGAAAAAAGgcatgcGAAGACGT*GTGAGCAAGGGCGAGGAGAATAACATG GCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGCTCCGTGAACGGCCAC GAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAA GCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTAACCCCCAACTTC ACCTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTC CTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGAC CGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGC ACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAAACCATGGGCTGGGAGGCCTCCT CCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGATGAGGCTGAAG CTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCC GTGCAGCTGCCCGGCGCCTACATCGTCGGCATCAAGTTGGACATCACCTCCCACAACGAGG ACTACACCATCGTGGAACTGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGG ACGAGCTGTACAAGTAG*AACCGGGAGCGCTGTGAATACAGTGCTCCCTTTTTTTATTGACG TC

The presented sequence is the modification made to the plasmid pCsrA-NYFP+2MS2bd. The underlined sequences are the cut sites for endonucleases SpeI, SphI, and AatII. The sequence was added to the plasmid between the SpeI and AatII sites. The italicized sequence is that of mStrawberry. All TriFC plasmids were created by inserting the sequence from Appendix A.1 between the cut sites for SpeI and SphI.

Table A.4 RBS modification to CsrA expression system

CsrA Expression System	Sequence Modification
pHL600	atatcccgggAGGAGGAAAAAAA**ATGCTGATTCTGACTCGTCG**
CsrA-200	atatcccgggAACCGCCCATAAGAATCCAAACAAAAGATCGCAT**ATGC TGATTCTGACTCGTCG**
CsrA-350	atatcccgggCCACAGTACAAATCGTTTAAGAGCACGAGCACATC**ATG CTGATTCTGACTCGTCG**
CsrA-750	atatcccgggATTACATAAGAGATACGTAAGCGACCTTAACGAT**ATGC TGATTCTGACTCGTCG**
CsrA-850	atatcccgggAATAAGAAGAAAGAACTTATTTCGGAGCTTTCAAA**ATG CTGATTCTGACTCGTCG**
CsrA-1500	atatcccgggAAAATCCTTCATATCCAATACAAGGCAAGAGATA**ATGC TGATTCTGACTCGTCG**
CsrA-2000	atatcccgggTCGTTTACATTAAGAAGCCCATAAGGCAGTTACT**ATGCT GATTCTGACTCGTCG**

This table represents the changes made to the CsrA expression system used in the titration assay. The original system uses the plasmid pHL600 (94) and is presented in the first entry. The expression system was modified by the insertion of the sequence designed by the RBS Calculator (104) between the XmaI cut site (cccggg) and the coding sequence for CsrA (bold). The putative RBS strength (arbitrary units) is given in the name of the system. The original pHL600 plasmid had a putative strength of 9000 (arbitrary units) as determined by the RBS Calculator.

References

1. Guillier M & Gottesman S (2006) Remodelling of the *Escherichia coli* outer membrane by two small regulatory RNAs. *Molecular Microbiology* 59(1):231-247.
2. Lenz D, Miller M, Zhu J, Kulkarni R, & Bassler B (2005) CsrA and three redundant small RNAs regulate quorum sensing in *Vibrio cholerae*. *Molecular Microbiology* 58(4):1186-1202.
3. Tu K & Bassler B (2007) Multiple small RNAs act additively to integrate sensory information and control quorum sensing in *Vibrio harveyi*. *Genes and Development* 21(2):221-233.
4. Romby P, Vandenesch F, & Wagner E (2006) The role of RNAs in the regulation of virulence-gene expression. *Current Opinion in Microbiology* 9(2):229-236.
5. Wassarman KM (2002) Small RNAs in bacteria: diverse regulators of gene expression in response to environmental changes. *Cell* 109(2):141-144.
6. Massé E & Gottesman S (2002) A small RNA regulates the expression of genes involved in iron metabolism in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* 99(7):4620-4625.
7. Huang H-YY*, et al.* (2008) sRNAMap: genomic maps for small non-coding RNAs, their regulators and their targets in microbial genomes. *Nucleic Acids Research* 37(Database issue):150-154.
8. Storz G, Vogel J, & Wassarman KM (2011) Regulation by small RNAs in bacteria: expanding frontiers. *Molecular Cell* 43(6):880-891.
9. Lioliou E, Romilly C, Romby P, & Fechter P (2010) RNA-mediated regulation in bacteria: from natural to artificial systems. *New Biotechnology* 27(3):222-235.
10. Wassarman KM, Zhang A, & Storz G (1999) Small RNAs in *Escherichia coli*. *Trends in Microbiology* 7(1):37-45.
11. Lanza AM, Blazeck JJ, Crook NC, & Alper HS (2012) Linking yeast gcn5p catalytic function and gene regulation using a quantitative, graded dominant mutant approach. *PLoS ONE* 7(4):e36193.
12. Glisovic T, Bachorik JL, Yong J, & Dreyfuss G (2008) RNA-binding proteins and post-transcriptional gene regulation. *FEBS Letters* 582(14):1977-1986.
13. Beisel CL & Storz G (2010) Base pairing small RNAs and their roles in global regulatory networks. *FEMS Microbiology Reviews* 34(5):866-882.
14. Vanderpool C, Balasubramanian D, & Lloyd C (2011) Dual-function RNA regulators in bacteria. *Biochimie* 93(11):1943-1949.
15. Martınez-Antonio A & Collado-Vides J (2003) Identifying global regulators in transcriptional regulatory networks in bacteria. *Current Opinion in Microbiology* (6):482-489.
16. Vakulskas CA, Potts AH, Babitzke P, Ahmer BMM, & Romeo T (2015) Regulation of bacterial virulence by Csr (Rsm) systems. *Microbiology and Molecular Biology Reviews* 79(2):193-224.

17. Romeo T, Gong M, Liu MY, & Brun-Zinkernagel AM (1993) Identification and molecular characterization of csrA, a pleiotropic gene from Escherichia coli that affects glycogen biosynthesis, gluconeogenesis, cell size, and surface properties. *Journal of Bacteriology* 175(15):4744-4755.
18. Weilbacher T*, et al.* (2003) A novel sRNA component of the carbon storage regulatory system of Escherichia coli. *Molecular Microbiology* 48(3):657-670.
19. Liu MY, Yang H, & Romeo T (1995) The product of the pleiotropic *Escherichia coli* gene *csrA* modulates glycogen biosynthesis via effects on mRNA stability. *Journal of Bacteriology* 177(10):2663-2672.
20. Timmermans J & Melderen VL (2009) Conditional essentiality of the csrA gene in *Escherichia coli*. *Journal of Bacteriology* 191(5):1722-1724.
21. McKee A*, et al.* (2012) Manipulation of the carbon storage regulator system for metabolite remodeling and biofuel production in *Escherichia coli*. *Microbial Cell Factories* 11(79):1-12.
22. Romeo T & Tatarko M (2001) Disruption of a global regulatory gene to enhance central carbon flux into phenylalanine biosynthesis in *Escherichia coli*. *Current Microbiology* 43(1):26-32.
23. Babitzke P & Romeo T (2007) CsrB sRNA family: sequestration of RNA-binding regulatory proteins. *Current Opinion in Microbiology* 10(2):156-163.
24. Edwards AN*, et al.* (2011) Circuitry linking the Csr and stringent response global regulatory systems. *Molecular Microbiology* 80(6):1561-1580.
25. Gudapaty S, Suzuki K, Wang X, Babitzke P, & Romeo T (2001) Regulatory interactions of Csr components: the RNA binding protein CsrA activates csrB transcription in *Escherichia coli*. *Journal of Bacteriology* 183(20):6017-6027.
26. Jonas K*, et al.* (2010) Complex regulatory network encompassing the Csr, c-di-GMP and motility systems of *Salmonella typhimurium*. *Environmental Microbiology* 12(2):524-540.
27. Romeo T, Vakulskas C, & Babitzke P (2013) Post-transcriptional regulation on a global scale: form and function of Csr/Rsm systems. *Environmental Microbiology* 15(2):313-324.
28. Timmermans J & Van Melderen L (2010) Post-transcriptional global regulation by CsrA in bacteria. *Cellular and Molecular Life Sciences : CMLS* 67(17):2897-2908.
29. Dubey AK, Baker CS, Romeo T, & Babitzke P (2005) RNA sequence and secondary structure participate in high-affinity CsrA-RNA interaction. *RNA* 11(10):1579-1587.
30. Liu M*, et al.* (1997) The RNA molecule CsrB binds to the global regulatory protein CsrA and antagonizes its activity in *Escherichia coli*. *The Journal of Biological Chemistry* 272(28):17502-17510.
31. Lapouge K*, et al.* (2013) RNA pentaloop structures as effective targets of regulators belonging to the RsmA/CsrA protein family. *RNA Biology* 10(6):1031-1041.

32. Mercante J, Edwards AN, Dubey AK, Babitzke P, & Romeo T (2009) Molecular geometry of CsrA (RsmA) binding to RNA and its implications for regulated expression. *Journal of Molecular Biology* 392(2):511-528.
33. Duss O, Michel E, Diarra dit Konte N, Schubert M, & Allain FH (2014) Molecular basis for the wide range of affinity found in Csr/Rsm protein-RNA recognition. *Nucleic Acids Research* 42(8):5332-5346.
34. Wang X*, et al.* (2005) CsrA post-transcriptionally represses pgaABCD, responsible for synthesis of a biofilm polysaccharide adhesin of *Escherichia coli*. *Molecular Microbiology* 56(6):1648-1663.
35. Dubey AK*, et al.* (2003) CsrA regulates translation of the Escherichia coli carbon starvation gene, cstA, by blocking ribosome access to the cstA transcript. *Journal of Bacteriology* 185(15):4450-4460.
36. Baker CS*, et al.* (2007) CsrA inhibits translation initiation of *Escherichia coli* hfq by binding to a single site overlapping the Shine-Dalgarno sequence. *Journal of Bacteriology* 189(15):5472-5481.
37. Yakhnin H*, et al.* (2011) CsrA represses translation of sdiA, which encodes the N-acylhomoserine-L-lactone receptor of *Escherichia coli*, by binding exclusively within the coding region of sdiA mRNA. *Journal of Bacteriology* 193(22):6162-6170.
38. Patterson-Fortin LM, Vakulskas CA, Yakhnin H, Babitzke P, & Romeo T (2013) Dual posttranscriptional regulation via a cofactor-responsive mRNA leader. *Journal of Molecular Biology* 425(19):3662-3677.
39. Yakhnin AV*, et al.* (2013) CsrA activates flhDC expression by protecting flhDC mRNA from RNase E mediated cleavage. *Molecular Microbiology* 87(4):851-866.
40. Park H, Yakhnin H, Connolly M, & Romeo T (2015) CsrA participates in a PNPase autoregulatory mechanism by selectively repressing translation of pnp transcripts that have been previously processed by RNase III. *Journal of Bacteriology* 197(24):3751-3759.
41. Holmqvist E, Wright PR, Li L, & Bischler T (2016) Global RNA recognition patterns of post-transcriptional regulators Hfq and CsrA revealed by UV crosslinking in vivo. *The EMBO Journal* 35(9):991-1011.
42. Kulkarni PR*, et al.* (2014) A sequence-based approach for prediction of CsrA/RsmA targets in bacteria with experimental validation in *Pseudomonas aeruginosa*. *Nucleic Acids Research* 42(11):6811-6825.
43. Jonas K*, et al.* (2008) The RNA binding protein CsrA controls cyclic di-GMP metabolism by directly regulating the expression of GGDEF proteins. *Molecular Microbiology* 70(1):236-257.
44. Edwards AN*, et al.* (2011) Circuitry linking the Csr and stringent response global regulatory systems. *Molecular Microbiology* 80(6):1561-1580.
45. Revelles O*, et al.* (2013) The Carbon Storage Regulator (Csr) system exerts a nutrient-specific control over central metabolism in *Escherichia coli s*train Nissle 1917. *PLoS ONE* 8(6):e66386.

46. Esquerré T*, et al.* (2016) The Csr system regulates genome-wide mRNA stability and transcription and thus gene expression in *Escherichia coli*. *Scientific Reports* 6:25057.
47. Tsai B, Wang X, Huang L, & Waterman M (2011) Quantitative profiling of in vivo-assembled RNA-protein complexes using a novel integrated proteomic approach. *Molecular and Cellular Proteomics : MCP* 10(4).
48. Barrick J, Sudarsan N, Weinberg Z, Ruzzo W, & Breaker R (2005) 6S RNA is a widespread regulator of eubacterial RNA polymerase that resembles an open promoter. *RNA* 11(5):774-784.
49. Wang Z, Gerstein M, & Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* 10:57-63.
50. Hafner M, Landthaler M, Burger L, & Khorshid M (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* (141):129-141.
51. Zhang C & Darnell RB (2011) Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nature Biotechnology* 29(7):607-615.
52. Kishore S, Jaskiewicz L, Burger L, & Hausser J (2011) A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. *Nature Methods* 8(7):559-564.
53. Gagnon KT & Maxwell ES (2011) Electrophoretic mobility shift assay for characterizing RNA–protein interaction. *RNA: Methods and Protocols* (703):275-291.
54. Stoltenburg R, Reinemann C, & Strehlitz B (2007) SELEX--a (r)evolutionary method to generate high-affinity nucleic acid ligands. *Biomolecular Engineering* 24(4):381-403.
55. Hall CE & Slayter HS (1959) Electron microscopy of ribonucleoprotein particles from *E. coli*. *Journal of Molecular Biology* 1:329-332.
56. Coloma R, Valpuesta JM, Arranz R, & Carrascosa JL (2009) The structure of a biologically active influenza virus ribonucleoprotein complex. *PLoS Pathogens* 5(5):e10000462.
57. Truong K & Ikura M (2001) The use of FRET imaging microscopy to detect protein–protein interactions and protein conformational changes in vivo. *Current Opinion in Structural Biology* (11):573-578.
58. Ciruela F (2008) Fluorescence-based methods in the study of protein–protein interactions in living cells. *Current Opinion in Biotechnology* (19):338-343.
59. Bernstein DS, Buter N, Stumpf C, & Wickens M (2002) Analyzing mRNA–protein complexes using a yeast three-hybrid system. *Methods* (26):123-141.
60. Kerppola TK (2007) Bimolecular fluorescence complementation (BiFC) analysis as a probe of protein interactions in living cells. *Annual Review of Biophysics* 37:465-487.

61. Valencia-Burton M, McCullough R, Cantor C, & Broude N (2007) RNA visualization in live bacterial cells using fluorescent protein complementation. *Nature Methods* 4(5):421-427.
62. Tyagi S (2007) Splitting or stacking fluorescent proteins to visualize mRNA in living cells. *Nature Methods* 4(5):391-392.
63. Kellermann S, Rath A, & Rentmeister A (2013) Tetramolecular fluorescence complementation for detection of specific RNAs in vitro. *Chembiochem : a European Journal of Chemical Biology* 14(2):200-204.
64. Rackham O & Brown CM (2004) Visualization of RNA-protein interactions in living cells: FMRP and IMP1 interact on mRNAs. *The EMBO Journal* 23(16):3346-3355.
65. Salgado H, Peralta-Gil M, & Gama-Castro S (2013) RegulonDB v8. 0: omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. *Nucleic Acids Research* 41:D203-D213.
66. Anne S, Katja NR, & Frank JB (2012) Transcription Stochasticity of Complex Gene Regulation Models. *Biophysical Journal* 103:1152-1161.
67. Ying X*, et al.* (2011) sTarPicker: A Method for Efficient Prediction of Bacterial sRNA Targets Based on a Two-Step Model for Hybridization. *PLoS ONE* 6(7):e22705.
68. Dias R & Kolazckowski B (2015) Different combinations of atomic interactions predict protein-small molecule and protein-DNA/RNA affinities with similar accuracy. *Proteins: Structure* (83):2100-2114.
69. Rozas I, Alkorta I, & Elguero J (2005) Modelling protein–RNA interactions: an electron density study of the guanidinium and formate complexes with RNA bases. *Organic and Biomolecular Chemistry* (3):366-371.
70. Zhou Q (2010) On weight matrix and free energy models for sequence motif detection. *Journal of Computational Biology* 17(12):1621-1638.
71. Granek JA & Clarke ND (2005) Explicit equilibrium modeling of transcription-factor binding and gene regulation. *Genome Biology* 6(10):R87.81-R87.10.
72. Roider HG, Kanhere A, Manke T, & Vingron M (2007) Predicting transcription factor affinities to DNA from a biophysical model. *Bioinformatics* 23(2):134-141.
73. Kollman PA, Massova I, Reyes C, & Kuhn B (2000) Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Accounts of Chemical Research* (33):889-897.
74. Mercante J, Suzuki K, Cheng X, Babitzke P, & Romeo T (2006) Comprehensive alanine-scanning mutagenesis of *Escherichia coli* CsrA defines two subdomains of critical functional importance. *The Journal of Biological Chemistry* 281(42):31832-31842.
75. Espah Borujeni A & Salis HM (2016) Translation initiation is controlled by RNA folding kinetics via a ribosome drafting mechanism. *Journal of the American Chemical Society* 138(22):7016-7023.
76. Shyu YJ & Hu C-DD (2008) Fluorescence complementation: an emerging tool for biological research. *Trends in Biotechnology* 26(11):622-630.

77. Yiu HW, Demidov VV, Toran P, Cantor CR, & Broude NE (2011) RNA detection in live bacterial cells using fluorescent protein complementation triggered by interaction of two RNA aptamers with two RNA-binding peptides. *Pharmaceuticals* (4):494-508.
78. Lowary P & Uhlenbeck O (1987) An RNA mutation that increases the affinity of an RNA-protein interaction. *Nucleic Acids Research* 15(24):10483-10493.
79. Peabody D & Ely K (1992) Control of translational repression by protein-protein interactions. *Nucleic Acids Research* 20(7):1649-1655.
80. Hu C-D, Chinenov Y, & Kerppola T (2002) Visualization of interactions among bZIP and Rel family proteins in living cells using bimolecular fluorescence complementation. *Molecular Cell* 9(4):789-798.
81. Contreras LM, Huang T, Piazza CL, Smith D, & Qu G (2013) Group II intron–ribosome association protects intron RNA from degradation. *RNA* 19:1497-1509.
82. Bardwell V & Wickens M (1990) Purification of RNA and RNA-protein complexes by an R17 coat protein affinity method. *Nucleic Acids Research* 18(22):6587-6594.
83. Keryer-Bibens C, Barreau C, & Osborne HB (2008) Tethering of proteins to RNAs by bacteriophage proteins. *Biology of the Cell /Uunder the Auspices of the European Cell Biology Organization* 100(2):125-138.
84. Romeo T (1998) Global regulation by the small RNA-binding protein CsrA and the non-coding RNA molecule CsrB. *Molecular microbiology* 29(6):1321-1330.
85. Chambers JR & Sauer K (2013) Small RNAs and their role in biofilm formation. *Trends in Microbiology* 21(1):39-49.
85. Jonas K & Melefors Ö (2009) The Escherichia coli CsrB and CsrC small RNAs are strongly induced during growth in nutrient-poor medium. *FEMS Microbiology Letters* 297(1):80-86.
87. Roederer M, Treister A, Moore W, & Herzenberg L (2001) Probability binning comparison: a metric for quantitating univariate distribution differences. *Cytometry* 45(1):37-46.
88. Liu J, Valencia-Sanchez M, Hannon G, & Parker R (2005) MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies. *Nature Cell Biology* 7(7):719-723.
89. Ozawa T, Natori Y, Sato M, & Umezawa Y (2007) Imaging dynamics of endogenous mitochondrial RNA in single living cells. *Nature Methods* 4(5):413-419.
90. Kostecki J, Li H, Turner R, & DeLisa M (2010) Visualizing interactions along the *Escherichia coli* twin-arginine translocation pathway using protein fragment complementation. *PloS ONE* 5(2):e9225.
91. Karlebach G & Shamir R (2008) Modelling and analysis of gene regulatory networks. *Nature Reviews Molecular Cell Biology* 9(10):770-780.
92. Endy D & Brent R (2001) Modelling cellular behaviour. *Nature* 409:391-395.
93. Adamson DN & Lim HN (2013) Rapid and robust signaling in the CsrA cascade via RNA–protein interactions and feedback regulation. *Proceedings of the*

National Academy of Sciences of the United States of America 110(32):13120-13125.
94. Komarova AV, Tchufistova LS, Dreyfus M, & Boni IV (2005) AU-rich sequences within 5 untranslated leaders enhance translation and stabilize mRNA in *Escherichia coli*. *Journal of Bacteriology* 187(4):1344-1349.
95. Barnard FM*, et al.* (2004) Global regulation of virulence and the stress response by CsrA in the highly adapted human gastric pathogen *Helicobacter pylori*. *Molecular Microbiology* 51(1):15-32.
96. Song H & Lee SY (2006) Production of succinic acid by bacterial fermentation. *Enzyme and Microbial Technology* 39:352-361.
97. Yakhnin H*, et al.* (2011) Complex regulation of the global regulatory gene csrA: CsrA-mediated translational repression, transcription from five promoters by Eσ70 and Eσ(S), and indirect transcriptional activation by CsrA. *Molecular microbiology* 81(3):689-704.
98. Gelderman G, Sivakumar A, Lipp S, & Contreras L (2015) Adaptation of Tri-molecular fluorescence complementation allows assaying of regulatory Csr RNA-protein interactions in bacteria. *Biotechnology and Bioengineering* 112(2):365-375.
99. Busch A, Richter AS, & Backofen R (2008) IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. *Bioinformatics* 24(24):2849-2856.
100. Cao Y, Zhao Y, Cha L, Ying X, & Wang L (2009) sRNATarget: a web server for prediction of bacterial sRNA targets. *Bioinformation* 3(9):364-2366.
101. Backofen R & Hess WR (2010) Computational prediction of sRNAs and their targets in bacteria. *RNA Biology* 7(1):33-42.
102. Campbell ZT, Bhimsaria D, & Valley CT (2012) Cooperativity in RNA-protein interactions: global analysis of RNA binding specificity. *Cell Reports* (1):570-581.
103. Salis HM, Mirsky EA, & Voigt CA (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nature Biotechnology* 27(10):946-953.
104. Zuker M & Stiegler P (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Research* 9(1):133-148.
105. Lorenz R & Bernhart SH (2011) ViennaRNA Package 2.0. *Algorithms for Molecular Biology* 6(26):1-12.
106. Espah Borujeni A, Channarasappa AS, & Salis HM (2014) Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. *Nucleic Acids Research* 42(4):2646-2659.
107. Hao Y*, et al.* (2011) Quantifying the sequence-function relation in gene silencing by bacterial small RNAs. *Proceedings of the National Academy of the Sciences of the Unites States fo America* 108(30):12473-12478.

108. Duss O, Michel E, Yulikov M, Schubert M, & Jeschke G (2014) Structural basis of the non-coding RNA RsmZ acting as a protein sponge. *Nature* 509:588-592.